

Responses to Coimisiún na Meán Call for Inputs: Online Safety Code

Publication Date: December 2023

Please Note: The responses collated in this document reference different types of online harms and age-inappropriate content. The submissions also include links to websites, research and other material intended to support the responses made to the Call for Inputs. Some readers may find some of the content distressing or upsetting. Therefore, reader discretion is advised.

List of Submissions:

1. NICAM
2. Alcohol Action Ireland
3. St. Louise's and St. Clare's Units, Children's Health Ireland
4. Women's Aid
5. Yoti
6. Carnegie UK
 - a. Carnegie UK Annex 1, Model Code A
7. Dublin City University Anti-Bullying Centre
8. Safe Ireland
9. Baby Formula Law Group Ireland
10. 5Rights Foundation
11. Age Verification Providers Association
12. Belong To LGBTQ+ Youth Ireland
13. Bodywhys – The Eating Disorders Association of Ireland
14. Competition and Consumer Protection Commission
15. Children's Rights Alliance
16. Conradh na Gaeilge
17. Ombudsman for Children's Office
18. Cybersafe Kids
19. Dairy Industry Ireland: IBEC
20. Department of Health
21. Food Drink Ireland: IBEC
22. FSM - Freiwillige Selbstkontrolle Multimedia-Diensteanbieter e.V.
23. Google Ireland Limited
24. Headline (with input from Shine and See Change)
25. Irish Heart Foundation
26. National Suicide Research Foundation
27. HSE National Office for Suicide Prevention
28. Meta Platforms Ireland Limited
29. Ofcom, UK
30. Spunout
31. Advertising Standards Authority Ireland

Coimisiún na Meán

- 32.** Irish Safer Internet Centre
 - a. ISIC App 1 Webwise Youth Panel
 - b. ISIC App 2 NPC Parents' Survey
 - c. ISIC App 3 NPC Parents and children's
 - d. ISIC App 4 NPC Children's survey
- 33.** Technology Ireland
- 34.** Dr. Susan Leavy and Dr. Ruihai Dong, University College Dublin
- 35.** VerifyMy
- 36.** WeProtect Global Alliance
- 37.** Commisariaat Vor de Media, Netherlands
- 38.** Irish Council for Civil Liberties
- 39.** The Internet Commission (Trust Alliance Group)
- 40.** Dr. Brian O'Neill, Emeritus Professor, TU Dublin
- 41.** Department of Children, Equality, Disability, Integration and Youth
- 42.** Ministry of the Interior and Kingdom Relations, Netherlands
- 43.** NewsBrands Ireland and Local Ireland
- 44.** National Women's Council of Ireland
- 45.** Rape Crisis Centre Managers Forum
- 46.** Anonymised Submission
- 47.** Samaritans Ireland
- 48.** Eurochild
- 49.** eSafety Commissioner
- 50.** TikTok
- 51.** National Parent's Council Primary
- 52.** Internet Watch Foundation
- 53.** Rape Crisis Network Ireland
- 54.** Irish Traveller Movement
- 55.** Data Protection Commission

Consultation questions and NICAM answers

Question 1: What do you think our main priorities and objectives should be in the first binding Online Safety Code for VSPS? What are the main online harms you would like to see it address and why?

The main objectives should be to keep children safe on VSPS from physical and mental dangers. Furthermore, banning illegal content and behaviour from VSPS and providing content information that children of all ages understand in the blink of an eye is crucial, in order to empower them to choose what they want to watch. Especially since our research shows that watching VSPS is mostly done individually by children without parental oversight. In the online harm they are facing we distinguish content and context.

Content rating

Based on international scientific research in the fields of child development, media psychology and communication science we know that content that might be harmful to children includes: violence, fear, sex, smoking, drugs and alcohol abuse, discrimination, coarse language, dangerous behaviour like challenges and stunts, suicide, self-harm and animal cruelty. These elements should be included and information on these elements should be taken into account when developing technical protection measures and provided to the viewers in easy to understand ratings such as pictograms, as written ratings can be difficult for young children to understand.

Generating reliable and independent content information on these elements forms the basis of any form of protection. Only based on this information can parental controls and age verification tools be effective protection measures on VSPS.

Generating this information can be done automatically and/ or by uploaders as long as it is based on uniform criteria that are applied across all VSPS. These criteria can be translated to an age recommendation and content advice. The criteria themselves should not be up to VSPS to choose but should be grounded in scientific research and theory.

This rating system can be administered by an independent body in the memberstate in which a certain VSP provider is located. Through the ERGA members the universal criteria can be discussed/ redefined and continuously developed based on scientific research into harmful effects on children.

This will ensure that, independent of the memberstate in which certain VSPS are registered, the same up to date criteria and ratings will apply.

We propose to join hands on this important topic and spread this message in Europe to make this the success that children deserve.

The number one priority is the content analysis (by the uploaders) of the video's, following uniform and independent standards to determine whether the content contains harmful elements. This is needed in order to protect and provide children and parents with reliable and trustworthy information on the content before watching. It is important that these criteria are uniform across platforms, so that parents and children understand what they can expect, thereby empowering parents and children to make their own decisions on what to watch and when to wait. In our research, children indicate that this is what they are currently missing on VSPS. Existing warnings are often vague and inconsistent, and it is unclear to children whom this information is coming from and whether they can be trusted.

When the content is analysed, the next step is age verification. Age verification systems and parental controls only work when universal criteria for video content analysis are being applied (and updated regularly), checked (and sanctioned) and explained to parents and children. Age verification for using the services should be based on national laws and when using the platforms, based on solid age and content ratings. Additionally mechanisms and content filtering suitable for different age groups should also be part of the Code.

Context

Online harms can be emphasized or generated by frequency of occurrence of topics in a child's media menu. This is controlled by algorithms, especially on the (shortform) VSPS that utilize automated feeds, suggestions and/ or recommendations based on interests or viewing history. Repetitive exposure to certain ideas, thoughts and/ or actions can generate harm by normalization. Eating disorders, self-harm, etc. are subject in which big risks can currently be seen.

This contextual algorithmic factor should be addressed in the code. Technical execution would vary from VSP to VSP but in general terms we would recommend to adopt an algorithm after consent approach. Meaning the platform is fit for all unless indicated that you are an adult (account) or have an adult's consent:

- Adults with an account consent to (use of) the service and only then are preferences stored and content that would be 12 or higher shown.
- Viewers without an account can only access the general content that does not contain any harmful elements mentioned above, and without any storage of/ acting on personal preferences.
- Children only should be able to create an account with parental consent. Once identified as a child, only self-indicated interests and preferences can be stored and these should be transparent to their parents. This will allow for additional protective measures and parental controls.

Question 2: What types of online harms do you think should attract the most stringent risk mitigation measures by VSPS?

How could we evaluate the impact of different types of harms e.g. severity, speed at which harm may be caused? Is there a way of classifying harmful content that you consider it would be useful for us to use?

NICAM has been working with a scientific approach towards standardized content analysis linked to international scientific research into developmental stages of children. This has led to a reliable system for determining until what age children could better not watch certain content. The objective, standardized analysis based on the actual content (what you see/ hear) in combination with the constant development of the criteria based on scientific research and media developments, offers a robust system for evaluating the different types of harm up to a certain age.

Real physical harm which can be caused by video's showing promotions or how to's on: suicide, auto mutilation, sexual abuse, dangerous challenges, drinking/ drug use, promotion of discrimination or violence, eating disorders, should be dealt with most strictly. The other content categories mentioned under question 1 (e.g., scary images, coarse language) follow.

Question 3: Do you have reports, academic studies or other relevant independent research that would support your views? If you do, please share them with us with links to relevant reports, studies or research.

We have performed research on shortform (Instagram & TikTok) focusing on the content as well as on the uploaders and users. The report is currently only available in Dutch, but we can present the results to you in English.

Furthermore, many recent academic studies have focused on the potential harms and need for better regulation of certain online content, for instance dangerous challenges (e.g., [Astorri et al., 2022](#)) the promotion of alcohol (e.g., [Hendriks et al., 2020](#)) and other substances (e.g., vaping) and eating disorders and other types of self-harm behaviours ([Harriger et al., 2022](#)).

With regard to the universality of ratings, academic research also shows detailed information, content-based ratings, and universal ratings are preferable for parents (see [Gentile et al., 2011](#)).

Question 4: What approach do you think we should take to the level of detail in the Code? What role could non-binding guidance play in supplementing the Code?

NICAM's experience with rules and regulation for industry partners is that the protection of minors is best served by a binding code.

Flexibility should be factored into the content of the code, not in the way parties should implement it.

A high level of detail and minimum standards for certain measures are necessary to safeguard children's rights online.

With regard to age verification, algorithms and ratings for instance, VSPS often have interests that conflict with the protection of minors. Furthermore, an objective (academic) basis is needed to determine the most appropriate age ratings for various content types, which is a necessary prerequisite for trustworthy ratings across platforms.

Question 5: What do you think would be the most effective structure for the Code? What are the most important factors we should consider when we decide how to structure the Code?

-

Question 6: How should we design the Code to minimise the potential for conflict and maximise the potential for synergies in how platforms comply with it and the DSA?

-

Question 7: To what extent, if at all, should the Code require VSPS providers to take measures to address content connected to video content?

Focus on the uniform analysis of video content on VSPS, except for where it touches on the risks mentioned in the first question. Automated analysis of connected contextual content could help.

Question 8: How should we ask VSPS providers to introduce a feature that allows users to declare when videos contain advertising or other type of commercial communications? Should the Code include specific requirements about the form in which the declaration should take? What current examples are there that you regard as best practice?

-

Question 9: How should we ask VSPS providers to introduce and design a flagging mechanism in the Code? How can we ensure that VSPS providers introduce the mechanism in a user-friendly and transparent way? How should we ask VSP Providers to report the decisions they've made on content after it has been flagged? To what extent should we align the Code with similar provisions on flagging in the DSA?

NICAM's experience with this is that an independent organization overseeing a flagging and/or complaints procedure, including the obligation to publish decisions made, is in the public interest and strengthens the reliability of the method/ system.

Within Kijkwijzer we are working with an independent complaints board to deal with complaints from the public. After a decision has been made by this board, it is published on our website for the public to read.

Question 10: What requirements should the Code include about age verification and age assurance? What sort of content should be shown by default to users who are logged out or in private browsing mode and whose age cannot be verified or assured? What evidence is there about the effectiveness of age estimation techniques? What current practices do you regard as best practice? Where accounts are not age verified should default privacy settings be used, should content default to universal content and should contact by others be more limited?

The VSPS should in principle be a safe place for children. This means that content rated higher than say 12 would not be freely accessible without an account. When content is not rated, it should not be accessible to kids (i.e., treated the same as content with the highest age rating). When profiles are not logged in, only content suitable for all ages should be accessible. For this approach, it is necessary that all content gets rated (e.g., by uploaders during the uploading process) and age verification measures are in place.

Automated systems for age estimation are thereby an unnecessary measure that makes the users again responsible and offers additional risks in relation to privacy and reliability.

Question 11: What requirements should the Code have in relation to content rating? What do you consider to be current best practice? What experiences have you had using content rating systems on platforms and do you think they have been effective? What steps could we ask VSPS to take to ensure content is rated accurately by users?

Answer to question 1 applies.

Kijkwijzer NICAM performed research on the content available on YouTube, TikTok and Instagram and what elements should be deemed potentially harmful.

We also asked children about their social media use, experiences with harmful content, and their wants and needs regarding age and content rating. Furthermore, we also interviewed uploaders, featuring questions like: what is feasible, what technical protection measures and ways of informing the public can be used. Based on this research we developed a special system for these uploaders with which they can rate their own content fast and simple; 'Kijkwijzer Online for YouTube'. Specific elements like for example 'dangerous challenges and stunts' were added to the system. In our analyses we found that these challenges are frequently present on social media and that they can pose risks/ dangers to children. Children do see these quite often and are worried about this content, as are their parents.

Additionally, we tested our prototype of Kijkwijzer Online among Industry parties on YouTube. Our findings with industry so far is that the Dutch uploaders are cooperative and willing to implement a rating system. They support the mission to protect children against potential harmful content. However they do mention that without the platforms facilitating the ratings to be built into the platform, is not possible to be fully compliant. The solution for this would be to embed the use of age ratings and content pictograms within YouTube and other video sharing platforms. Hereby allowing uploaders to show the age and content ratings on the platform next to the title of a production as well as embedding them during the first 5 seconds in their video on a 'ratings layer'.

Therefore, we hereby request for CnM to include the obligation in the act for VSP's to facilitate (national) rating systems on their platforms by providing their uploaders with options to embed and show ratings on their platform(s).

Question 12: What requirements should the Code have in relation to parental control features? How can we ensure that VSPS providers introduce the mechanism in a user-friendly and transparent way? Can you point to any existing example of best practice in this area? Should parental controls be 'turned-on' by default for accounts of minors or where age is not verified?

Parental controls should be based on solid content ratings as described above. Only after logging into an account certain content will be available. **Harmful content should be turned off by default.**

Question 13: What requirements should the Code contain to ensure that VSPS provide for effective media literacy measures and tools?

The content ratings and local organization overseeing the system, should be clearly visible and pictograms should be explained to kids and parents in campaigns on the platform. It should be clear for parents and children what the content ratings mean, but also what they are based on and where they can file their complaints or gain more information about content ratings and media literacy measures / organizations.

Question 14: How should we ask VSPS providers to address online harms in their terms and conditions in the Code, including the harms addressed under Article 28b? How should key aspects of terms and conditions be brought to users' attention? What examples are there of best practice in relation to terms and conditions including content moderation policies and guidelines?

-

Question 15: How should we ask VSPS providers to address content moderation in the Code? Are there any current practices which you consider to be best practice? How should we address automated content detection and moderation in the Code?

Moderation on illegal content as mentioned in the first question, we recommend not moderation but information and technical protection measures for all the other content.

Question 16: What requirements should the Code include about procedures for complaint-handling and resolution, including out-of-court redress or alternative-dispute resolution processes? To what extent should these requirements align with similar requirements in the DSA? What current practices could be regarded as best practice? How frequently should VSPS providers be obliged to report to the Commission on their complaint handling systems and what should those reports contain? Should there be a maximum time-period for VSPS providers to handle user complaints and if so, what should that period be?

Independent complaints procedure safeguarding the public interest. We refer to question 9.

Question 17: What approach do you think the Code should take to ensuring that the safety measures we ask VSPS providers to take are accessible to people with disabilities?

-

Question 18: What approach do you think the Code should take to risk assessments and safety by design? Are there any examples you can point us towards which you consider to be best practice?

We refer to our approach to the previous questions. Basically, VSPS should be safe unless you are a registered user. Content should be deemed harmful, unless it has been awarded a solid age rating. With regard to design, it is important to be transparent about the platforms' algorithms. For children, highly personalized algorithms should not be used at all as it poses too many risks for this vulnerable group.

(See: <https://www.theguardian.com/technology/2023/apr/04/how-tiktoks-algorithm-exploits-the-vulnerability-of-children>).

Question 19: How do you think that cooperation with other regulators and bodies can help us to implement the Code for VSPS?

We refer to question 1. We think that ERGA could play a good role in the coordination of an international approach to content rating and information on VSPS.

Our outlook on this is that an international system for rating productions on VSP's should be implemented. In this system each memberstate could participate by which the protection of minors becomes universal and independent of the country in which a VSP is registered, creating a levelled playing field and solid protection for minors online.

Question 20: What approach do you think we should take in the Code to address feeds which cause harm because of the aggregate impact of the content they provide access to? Are there current practices which you consider to be best practice in this regard?

We refer to our answer to question 1 and more specifically 'context'.

In general terms this kind of recommendation should not be offered to children. And should not be allowed for adults unless you have a registered profile in which you accept this aggregation.

Question 21: Do you have any views on how requirements for commercial content arranged by a VSPS provider itself should be reflected in the Code?

-

Question 22: What compliance monitoring and reporting arrangements should we include in the Code?

-

Question 23: Should the Code have a transition period or transition periods for specific issues? Which areas touched on in this Call for Inputs may VSPS providers require time to transition the most? What time frame would be reasonable for a transition period?

The code should be implemented as soon as possible. The platforms already have systems in place for displaying content warnings, age verification etc. this should enable a smooth transition towards better and more uniform regulation.

Additional call for input from Children

Call for Inputs – How you can help? We are inviting organisations representing children and young people under the age of 25 to ask, where practicable, their members, including youth panels and youth committees, about the issues raised in our Call for Inputs. We have drafted several questions below that children and youth-focused groups and organisations may find useful for this. 27 These questions are indicative and you should feel free to tailor them to those that you represent or work with. We would be grateful if responses could be submitted together with your main response. Indicative Questions – Questions are about video-sharing platforms only.

Q1. What do you like about being able to watch or share videos on websites or apps?

Q2. How safe do you feel when you are watching or sharing videos on websites or apps?

Q3. Are you concerned about any videos that you see on websites or on apps? If you are, what types of videos concern you the most?

Q4. Do you feel that you have enough control over the type of videos that you see on websites or apps?

Q5. Do you think that companies who run websites or apps that allow videos to be watched or shared should do anything to make things safer for you or your friends or family?

Q6. How old do you think a child should be before they should be allowed to watch or share videos on websites or in apps? Should there be different rules for children who are different ages?

Q7. Have you ever reported your concerns to your parent/s or guardian/s or to a company in charge of websites or apps about a video that you have seen? How did that go?

Q8. Is there anything else you would like to comment on?

Relating NICAM Research

We recently conducted a study among Dutch children (through their parents) and teenagers to better understand their experiences with (potentially harmful content on) video sharing platforms, and their wants and needs regarding rating systems. Several years ago, we already did this for YouTube and now we focused on short form content (i.e., Tiktok and Instagram). The results show that more than 80% of Dutch teenagers between the ages of 10 and 16 and their parents would like warnings before shocking images on social media such as TikTok and Instagram. It concerns images with violence, sex, animal suffering or 'scary' things.. Young people are afraid that the algorithm will serve them more and more videos that will make them feel afraid, embarrassed or unsafe. They not only want a warning that videos contain shocking images, but also what kind of images, so that they can decide for themselves whether they would like to watch or not.

Almost one in five teenagers encounters nasty or shocking videos online. This worries parents: more than two-thirds are afraid that their child will see violent images, dangerous or bullying behaviour or sexual acts. Young people themselves find it unpleasant when shocking images are shown unannounced. At the moment, they are sometimes warned with banners on the platform saying things like 'contains sensitive content'. However, children find these too vague, unclear who the sender is, they do not really stand out and/ or the notification is incorrect.

Deciding for themselves

The research shows that 64% of teenagers (10-16 years old) watch videos on TikTok (for Instagram this is 39%). As they get older, they spend more time on these platforms. More than two-thirds watch TikTok for more than half an hour every day. They mainly watch vloggers and influencers, music, and game videos. Young people over the age of 12 decide for themselves what they do or do not watch, parents hardly watch together with them. Where in most cases there are rules and/ or agreements with young children about how long and what can be watched, for teenagers this is only the case for a small minority.

Difficult to handle

Since 1 July 2022, uploaders of online videos, who are based in the Netherlands, have a Chamber of Commerce registration, have more than 500,000 followers, and publish a minimum of 24 videos per year, must apply Kijkwijzer and warn for potentially harmful images for children. The research shows that the most shocking or violent videos are mainly uploaded by uploaders with smaller numbers of followers or uploaders who are not located in the Netherlands (or Europe). As a result, they do not have to comply to the Dutch law and providing Kijkwijzer information is therefore not mandatory for them. However, children and teenagers do get to see these videos.

In addition, videos on Instagram and TikTok are often posted immediately on the platform. As a result, it is not always possible to assess in advance and assign pictograms (warnings) to them, as is the case with the current application of Kijkwijzer. That is why NICAM is now working together with uploaders to see how we could make this work.

The research was conducted in collaboration with GFK and YoungWorks and consisted of interviews with children, young people, uploaders and analysis of content on social media.

Letter sent to Coimisiún na Meán

Situation in the Netherlands

The Ministry of Education, Culture and Science and the Dutch Media Authority and NICAM share the responsibility for the protection of minors within the media landscape in the Netherlands.

One independent body to safeguard children's interests in the protection of minors. The Netherlands Institute for the Classification of Audio-visual Media (NICAM) is a foundation/non-profit organization with one clear mission: protecting children from potentially harmful audio-visual content.

The word harmful is essential in this mission, since we do not judge on suitability or good taste of media. We base our advice on scientific research into potential harmful effects up to a certain age and on research among parents and children. The scientific committee monitors relevant international research and we perform (European) public consultations to see if we are performing well and research states that parents value the system as an important tool. This results in Kijkwijzer, a dynamic rating system.

The AVMS Directive was implemented into Dutch law in 2020. This changed the rules applicable to Video on Demand (VOD) services and Video Sharing Platforms (VSP's). Where VOD services were often already registered with NICAM, the uploaders on VSP's now also have to do so under the new law. Within the operationalization of the law, the Dutch Media Authority concluded that uploaders on YouTube, Instagram and TikTok, meeting certain criteria, are being viewed as Commercial Media Services (on Demand).

NICAM performed research on the content available on YouTube, TikTok and Instagram and what elements should be deemed potentially harmful. We also asked children about their use, needs and effects of social media and performed a qualitative study on uploaders. This research featured questions like: what is feasible, what technical protection measures and ways of informing the public can be used. Based on this research we developed a special system for these uploaders with which they can rate their own content fast and simple; 'Kijkwijzer Online for YouTube'. Specific elements like for example 'dangerous challenges and stunts' were added to the system. It was discovered within the content analysis that these challenges can pose risks/dangers to children. Children do see these quite often and are worried about this content. For short-form content that can be found on platforms like Instagram and TikTok we are currently researching alternatives, more in line with the quick and automated nature of these platforms.

Additionally, we tested our prototype of Kijkwijzer Online under industry parties on YouTube. Our findings with industry so far is that the Dutch uploaders are cooperative and willing to implement a rating system. They support the mission to protect children against potential harmful content. However they do mention that without the platforms facilitating the ratings to be built into the platform, it is not possible to be fully compliant. The solution for this would be to embed the use of age ratings and content pictograms within YouTube. Hereby allowing uploaders to show the age and content ratings on the platform next to the title of a production as well as embedding them during the first 5 seconds in their video on a 'ratings layer'.

Therefore, we hereby request for CnM to include the obligation in the act for VSP's to facilitate (national) ratings systems on their platforms by providing their uploaders with options to embed and show ratings on their platform(s).

Our outlook on this is that an international system for rating productions on VSP's should be implemented. In this system each memberstate could participate by which the protection of minors becomes universal.

The proposed display methods should enable uploaders to better inform children by being able to show and change ratings in case of a mistake or complaint. It is this independent rating information on which technical protection measures should be built like parental controls, kids profiles, responsible algorithms, etc. Therefore we need to make sure that the implementation of this information on VSP's is being covered now and for future VSP's independent of the EU country in which the VSP registers.

In this way a levelled playing field arises for uploaders throughout Europe, which is important for the support of uploaders and industry.

Additionally, the platform's cooperation is necessary for the adoption and understanding of the ratings on their platforms. A universal approach to this is of the essence for providing the information and create clarity and uniformity with both the public and the uploaders. Good protection starts with good information. Children have the right (UN) to reliable (independent) content information on every platform.

To summarize:

1. We hereby request to include in your act the obligation to facilitate the display of national rating systems on VSP's.
2. We propose to initiate 1 European rating system for VSP's in which we can take the lead together and form a working group with other member states.

For more information, please do not hesitate to contact us. We are very willing to provide you all necessary information and continue our conversation on this.



AlcoholAction
Ireland

Alcohol Action Ireland Submission to Coimisiún na Meán

Call For Inputs on Developing First Online Safety Code

Alcohol Action Ireland (AAI) was established in 2003 and is the national independent advocate for reducing alcohol harm. We campaign for the burden of alcohol harm to be lifted from the individual, community and State, and have a strong track record in campaigning, advocacy, research and information provision. Our work involves providing information on alcohol-related issues, creating awareness of alcohol-related harm and offering policy solutions with the potential to reduce that harm, with a particular emphasis on the implementation of the Public Health (Alcohol) Act 2018. Our overarching goal is to achieve a reduction in consumption of alcohol and the consequent health and social harms which alcohol causes in society.

Alcohol Action Ireland
Coleraine House
Coleraine Street
Dublin, D07 E8XF

Tel +353 1 878 0610 : admin@alcoholactionireland.ie: alcoholireland.ie

Alcohol Action Ireland CEO: Dr Sheila Gilheany; Directors: Catherine Brogan, Pat Cahill, Paddy Creedon, Michael Foy, Prof Jo-Hannah Ivers, Marie-Claire McAleer, Prof Frank Murray (Chair); Dr Colin O'Driscoll, Dr Mary O'Mahony, Dr Bobby Smyth.

Patron Prof. Geoffrey Shannon

Alcohol Action Ireland is a registered Irish Charity. Registered Charity Number: 20052713 Company No: 378738. CHY: 15342.

1.0 Introduction

Alcohol is one of the most heavily marketed products in our retail environment with the annual spend on alcohol marketing conservatively estimated at €115m in Ireland alone. Young people are an important market for the alcohol industry.

Comprehensive research now clearly tells us that alcohol marketing increases the likelihood that adolescents will start to use alcohol, and to drink more if they are already using alcohol.

Given that we know alcohol is no ordinary commodity but one that has been identified as one of the four industries (tobacco, unhealthy food, fossil fuel, and alcohol) responsible for at least a third of global deaths per year, we must protect children from it and its predatory marketing practices. The Joint Committee on Tourism, Culture, Arts, Sport and Media in its report on pre-legislative scrutiny of the general scheme of the Online Safety and Media Regulation Bill 2022, recommended a ban on advertising to children online, including, at the very minimum, advertising of junk foods, alcohol, foods high in fat, salt or sugar, and gambling.

It is clear from all of the available evidence that we cannot continue to allow big business to commodify childhood by allowing young people to fall under the influence of advertisers selling products detrimental to their health and well-being- or what global children's rights experts call exploitive marketing of unhealthy commodities and "an important threat to children's health and futures".

2.0 Protection from predatory commercial practices

Digital advertising is far more harmful for children than any other form. Estimates suggest that by the time a child turns 13, advertisers already hold over 72 million data points about him/her, and the surveillance advertising industry for children is worth in excess of \$1 billion. A recent World Health Organization report, noted that: "Alcohol marketing is adapting to new realities faster than current legal regulations across the region, with industry using opportunities offered by digital platforms to sell their products in a largely unregulated market". It also contained a stark warning on "the targeting of consumers including children and adolescents to promote drinking."

International experts on children's health and rights have also warned that "large companies incorporate the science of the life course approach into their marketing, to achieve the adherence and fidelity of children to capture future consumption".

Advertising restrictions have been assessed as highly cost-effective because they can influence the initiation of alcohol use and risk behaviour at the population level. Currently, Irish law- the Public Health Alcohol Act (section 14) prohibits the advertising of alcohol in certain public spaces with the aim of reducing the amount of advertising children see, but does not specify similar restrictions in the online environment. By allowing this gap to remain, in essence we are saying that in the real world children must be 200 metres from alcohol ads, but in the digital world, they have it in the palm of their hand 24/7. This is an irrational position. Furthermore, any gain made may have been undermined by the proliferation of the marketing of zero alcohol drinks, using the same branding as the master brand. This brand sharing is a real threat to the spirit and intention of the PHAA and AAI has been advocating to ensure advertising alcohol brands in areas protected by PHAA does not continue.

Alcohol brands are steadily spending more of their advertising budgets on digital marketing, likely because there's a younger audience and very little in the way of regulation. This makes keeping pace with the ever-evolving digital marketplace capturing the influence of our children even more urgent. Currently there is a Children's Commercial Communications Code which was devised by the Broadcasting Authority of Ireland. It is of note that the Statutory Report of the effect of this Code found that Diageo was the number four broadcast advertiser to children in Ireland. This is clearly highly unsatisfactory and illustrates that despite protestations from alcohol producers that they do not target children for their advertising, none the less children are seeing and absorbing their marketing.

It is essential that any new codes which are developed by Coimisiún na Meán must ensure that children are not targeted by alcohol advertisers either in online or traditional broadcast marketing.

The main priority and objective of the first binding Online Safety Code should be the protection of children and young people online. The UN Committee on the Rights of the Child are clear that ‘the rights of every child must be respected, protected and fulfilled in the digital environment.’ Further they recommend that ‘in all actions regarding the provision, regulation, design, management and use of the digital environment, the best interests of every child is a primary consideration. It follows then that the wellbeing of children must be given primacy over the commercial interests of the alcohol industry.

1. AAI recommends that the commission develops Codes that protect children and the general public from harmful commercial practices of the alcohol industry - including advertising to children (in spaces they inhabit), sponsorship and product placement.
2. The onus should be on alcohol producers and advertisers to provide evidence that their advertisements are not reaching children.
3. There should be an easily accessed facility which allows individuals to set controls so that they do not see alcohol advertisements if that is their choice.
4. Alcohol brands (including zero alcohol products) should not be allowed to sponsor programmes that are before 9pm and might be seen by children.
5. Alcohol brands should not be allowed to use sponsored content campaigns to reach young people across digital platforms and mediums, allowing them to normalise the visibility of a harmful product and drive consumption of alcohol.
6. The United Nations Committee on the Rights of the Child (UNCRC), new General Comment pertaining to the rights of the child in relation to the digital environment states that “States parties should prohibit by law the profiling or targeting of children of any age for commercial purposes on the basis of a digital record of their actual or inferred characteristics, including group or collective data, targeting by association or affinity”. Alcohol brands must be prohibited from collecting data on young people and must provide evidence that they are not tracking and profiling young people.

7. AAI welcomes the establishment of a Youth Advisory Committee to advise the Commission on its online safety work and in conducting research on online harms. The World Health Organisation in a 2021 [report on digital marketing of alcohol](#) have called for research to include children and young people's exposure to and engagement with digital marketing of alcohol. The impact of alcohol on children's lives has emerged as a consistent theme throughout consultations with them through the structure of the Comhairle na nÓg over the years. Additionally, the [2016, Children Seen and Heard report](#), younger children (8–12 years) mentioned alcohol abuse in the top four categories of things they disliked about Ireland. AAI's believes that by consulting with young people, we can better understand what needs to be achieved to fulfil children's rights obligations and protect children from alcohol harm and the predatory commercial practices of the alcohol industry.



AlcoholAction
Ireland



Online Safety – Developing Ireland’s First Binding Online Safety Code for Video-Sharing Platform Services

Submission by St Louise’s and St Clare’s Units, Children’s Health Ireland, Specialist Child Sexual Abuse Services (The Alders Unit)

Eimear Lacey, Principal Social Worker, St. Louise’s Unit

Rosaleen McElvaney, Principal Psychotherapist, St. Clare’s Unit

Christopher Behan, Senior Social Worker, St. Louise’s Unit

Oriel Smith, Senior Social Worker, St. Clare’s Unit

31 AUGUST 2023

1. Introduction:

St. Clare's and St. Louise's Units (SCU/SLU) [*soon to be renamed The Alders Unit*] welcome the call for inputs to inform the development of Ireland's first binding online safety code. These Units, based respectively in Children's Health Ireland in Tallaght and Blanchardstown, offer specialist sexual abuse service to children and young people from the ages of 3-18 years and their families. The catchment areas covered by these services are North Dublin City and County, Louth, Meath, Cavan and Monaghan (SCU/The Alders Unit at Blanchardstown) and South Dublin City and County, Wicklow and Kildare (SLU/The Alders Unit at Tallaght). The units offer a number of specialist services, including comprehensive assessment of child sexual abuse and exploitation concerns. Therapeutic interventions to children and families and consultation services to professionals regarding issues related to CSA concerns. The units have been in existence since 1988 and have extensive experience in working with children and families impacted by CSA. Over the past number of years, we have witnessed the increasing impact of the online space on the children and families we work with and acknowledge that there is a pressing need for a robust, unambiguous and binding code to protect people from online harms and to ensure that measures to address these harms are effective. We support the efforts of the Commission in ensuring that video sharing platforms take responsibility for protecting children from being able to access videos with violent, abusive or criminal content, bringing Irish society in line with Article 25 of the Directive 2011/93/EU of the European Parliament and of the Council of 13 December 2011 on combating the sexual abuse and sexual exploitation of children and child pornography (<https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A02011L0093-20111217>), the United Nations Charter on the Rights of the Child (https://www.ohchr.org/en/instruments_mechanisms/instruments/convention-rights-child), and the United Nations guiding principles on how businesses can play their part in child protection (https://www.ohchr.org/sites/default/files/documents/publications/guidingprinciplesbusinesshr_en.pdf). Material with violent, abusive or criminal content should not be accessible for children and it is the responsibility of society to ensure that children are protected from such material.

We have responded to the specific questions posed by Coimisiún na Meán below as succinctly as possible.

1) **Question 1: What do you think our main priorities and objectives should be in the first binding Online Safety Code for VSPS? What are the main online harms you would like to see it address and why?**

The following areas are those we have identified as requiring prioritisation in the first binding Online Safety Code for VSPS.

- Exposure to harmful sexual content

- CSAM dissemination and removal
- Grooming
- Age regulations
- Bullying and harassment
- Algorithmic targeting (content) – self-harm/suicide/eating disorders/CSAM/inappropriate adult content
- Exploitation

The impact and harm caused by the issues identified above are potentially long term and in some instances catastrophic to those who experience, are exposed to, or are actively engaged in creating content related to the above areas.

2) Question 2: What types of online harms do you think should attract the most stringent risk mitigation measures by VSPS? How could we evaluate the impact of different types of harms e.g. severity, speed at which harm may be caused? Is there a way of classifying harmful content that you consider it would be useful for us to use?

We have identified three areas we view as requiring stringent risk mitigation measures:

- I. CSAM – not removing and allowing sharing
- II. Evidence of grooming – e.g. attempts to move to encrypted platform such as WhatsApp etc.
- III. Content that encourages self-harm, suicidal ideation, eating disorders, and clearly presents a risk of harm to individuals’ mental health

In order to evaluate the impact of these harms we recommend that the Commission creates links with organisations who work with vulnerable populations to seek quantitative/qualitative information from them. This also represents an opportunity to create research partnerships to broaden knowledge, expertise and ultimately build effective responses to harms posed in this space. We are also of the view that the Commission should create a mechanism that will support individuals to self-report the impact of harms that they have experienced and also for professionals to report trends that they experience without the need for formal research which can take significant time before formal publication.

Classification: We recommend consideration of the following classifications of harmful content (these are not listed in order of priority).

- Mental health
- Sexually harmful/abusive material
- Adult sexual material
- Bullying and harassment
- Disinformation
- Extreme violence and aggression

3) Question 3: Do you have reports, academic studies or other relevant independent research that would support your views? If you do, please share them with us with links to relevant reports, studies or research.

- Caffo, E. (ed.) (2021). Online child sexual exploitation: Treatment and prevention of abuse in a digital world. Springer.
- Children's Commissioner (2018). *Life in 'likes': children's commissioner report into social media use among 8-12 year olds*.
- Chiu, J., & Quayle, E. (2022). Understanding online grooming: An interpretative phenomenological analysis of adolescents' offline meetings with adult perpetrators. *Child Abuse and Neglect*, 128, [105600]. <https://doi.org/10.1016/j.chiabu.2022.105600>
- Cooper, K., Quayle, E., Jonsson, L. & Svedin, C. G., (2016), Adolescents and self-taken sexual images: A review of the literature, *Computers in Human Behavior*, 55, 706-716
- CyberSafeKids. (2022). *Academic Year in Review 2021 – 2022*, https://www.cybersafekids.ie/wp-content/uploads/2022/09/CSK_YearInReview_2021-2022_FINAL.pdf
- Dragiewicz, M., Burgess, J., Matamoros-Fernández, A., Salter, M., Suzor, N. P., Woodlock, D., & Harris, B. (2018). Technology facilitated coercive control: Domestic violence and the competing roles of digital media plat forms, *Feminist Media Studies*, 18, 609 - 625, <http://dx.doi.org/10.1080/14680777.2018.1447341>
- Dragiewicz, M., Woodlock, D., Salter, M., & Harris, B. (2022). “What’s mum’s password?”: Australian mothers’ perceptions of children’s involvement in technology-facilitated coercive control, *Journal of Family Violence*, 37, 137 - 149, <http://dx.doi.org/10.1007/s10896-021-00283-4>
- Guerra, C., Pinto-Cortez, C., Toro, E., Efthymiadou, E., & Quayle, E. (2021). Online sexual harassment and depression in Chilean adolescents: Variations based on gender and age of the offenders. *Child Abuse and Neglect*, 120, [105219]. <https://doi.org/10.1016/j.chiabu.2021.105219>
- Hamilton-Giachritsis, C., Hanson, E., Whittle, H. & Beech, A. (2017). *"Everyone deserves to be happy and safe" A mixed methods study exploring how online and offline child sexual abuse impact young people and how professionals respond to it*. [10.13140/RG.2.2.35001.88164](https://doi.org/10.13140/RG.2.2.35001.88164).
- INHOPE (undated). CSAM media guidelines, INHOPE. <https://inhope.org/media/pages/support-us/media-guidelines/55949beee7-1644491683/csammediaguidelines.pdf>. Accessed on 29th August, 2023.
- Itzin, C. (2001). Incest, paedophilia, pornography and prostitution: Making familial abusers more visible as the abusers. *Child Abuse Review*, 10, 35–48.
- Jones, C., Salter, M., & Woodlock, D. (2023). “Someone who has been in my shoes”: The effectiveness of a peer support model for providing support to partners, family and friends of child sexual abuse material offenders, *Victims and Offenders*, 18, 715 - 731, <http://dx.doi.org/10.1080/15564886.2022.2051108>
- Jones, L. M., Mitchell, K. J., & Finkelhor, D. (2013). Online harassment in context: Trends from three youth internet safety surveys (2000, 2005, 2010). *Psychology of Violence*, 3(1), 53–69.

- Mäenpää, M., Ovaska, A., Santtila, P. & Korkman, J. (2022). *Analysis of current practices and identification of training gaps and needs of target groups: Ensuring child-friendly justice through the effective operation of the Barnahus Units in Finland*. Finnish Institute for Health & Welfare
- Livingstone, S., Kirwil, L., Ponte, C. & Staksrud, E. (2013). *In their own words: what bothers children online? with the EU Kids Online Network*. EU Kids Online, London School of Economics & Political Science, London, UK.
- <https://oneintenpodcast.org/episodes/growing-up-online-addressing-child-sex-tortion/>
- Palmer, T. (2015). *Digital dangers: The impact of technology on the sexual abuse and exploitation of children and young people*. Barnardos, Head Office, Essex.
- Quayle, E. (2016). *Researching online child sexual exploitation and abuse: Are there links between online and offline vulnerabilities?* . The London School of Economics and Political Science.
- Quayle, E. (2020). Prevention, disruption and deterrence of online child sexual exploitation and abuse. *ERA Forum*. <https://doi.org/10.1007/s12027-020-00625-7>
- Quayle, E. (2020). Online sexual deviance, pornography and child sexual exploitation material. *Forensische Psychiatrie, Psychologie, Kriminologie*, 14(3), 251-258. <https://doi.org/10.1007/s11757-020-00607-y>
- Quayle, E. (2022). Self-produced images, sexting, coercion and children's rights. *ERA Forum*, 23(2), 237-251. <https://doi.org/10.1007/s12027-022-00714-9>
- Quayle, E., & Cariola, L. (2017). *Youth-produced sexual images: A victim-centred consensus approach*. University of Edinburgh.
- Quayle, E., & Cariola, L. (2019). Management of non-consensually shared youth-produced sexual images: A Delphi study with adolescents as experts. *Child Abuse and Neglect*, 95, [104064]. <https://doi.org/10.1016/j.chiabu.2019.104064>
- Quayle, E. & Cooper, K., (2015). The role of child sexual abuse images in coercive and non-coercive relationships with adolescents: A thematic review of the literature, *Child and Youth Services*. 36(4) 312-328
- Quayle, E., Cooper, K., Newman, E., & Cariola, L. (2017). Deterrents to viewing indecent online images of children: A meta-narrative review. In *PROSPERO International prospective register of systematic reviews* (pp. 1-4). https://www.crd.york.ac.uk/PROSPERO/display_record.php?ID=CRD42017067498
- Quayle, E., Jonsson, L. S., Cooper, K., Traynor, J. & Svedin, C. G., (2018). Children in identified sexual images - who are they? Self- and non-self-taken images in the International Child Sexual Exploitation Image Database 2006-15, *Child Abuse Review*. 27(3), 223-238.
- Quayle, E., & Koukopoulos, N. (2019). Deterrence of online child sexual abuse and exploitation. *Policing: Journal of Policy and Practice*, 13(3), 345-362. <https://doi.org/10.1093/>
- Quayle, E., & Newman, E. (2016). An exploratory study of public reports to investigate patterns and themes of requests for sexual images of minors online. *Crime Science*, 5(1), [2]. <https://doi.org/10.1186/s40163-016-0050-0>
- Quayle, E., Schwannauer, M., Varese, F., Cartwright, K., Hewins, W., Chan, C., Newton-Braithwaite, A., Chitsabesan, P., Richards, C. & Bucci, S., (2023). The experiences of

- practitioners working with young people exposed to online sexual abuse, *Frontiers in Psychiatry*, 14, 1-14 <https://www.doi.org/10.3389/fpsyt.2023.1089888>
- Richardson, L. (11 July 2023) *Claims by academic group calling for EU to abandon CSAM-blocking policies don't stand up to real-world scrutiny*, Canadian Centre for Child Protection Blog, https://www.protectchildren.ca/en/press-and-media/blog/2023/EU_academic_letter_response
- Salter, M. (2016). Privates in the online public: Sex(ting) and reputation on social media, *New Media and Society*, 18, 2723 - 2739, <https://dx.doi.org/10.1177/1461444815604133>
- Salter, M. (2023). Online child sexual exploitation in the news: Competing claims of gendered and sexual harm. In K. Boyle, & S. Berridge, (eds.), *The Routledge companion on gender media violence*, Routledge.
- Salter, M. (2018). From geek masculinity to gamergate: The technological rationality of online abuse', *Crime, Media, Culture*, 14, 247 – 264. <https://dx.doi.org/10.1177/1741659017690893>
- Salter, M. & Hanson, E. (2021), "'I need you all to understand how pervasive this issue is": User efforts to regulate child sexual offending on social media", In J. Bailey, A. Flynn, A., and N. Henry (Eds.) *The emerald international handbook of technology-facilitated violence and abuse* (pp. 729-748), Emerald Publishing Limited. <https://doi.org/10.1108/978-1-83982-848-520211053>
- Salter, M., & Richardson, L. (2021). The Trichan takedown: Lessons in the governance and regulation of child sexual abuse material, *Policy and Internet*, 13, 385 – 399. <https://dx.doi.org/10.1002/poi3.256>
- Salter, M. & Sokolov, S. (2023) "Talk to strangers!": Omegle and the political economy of technology-facilitated child sexual exploitation, *Journal of Criminology*, Available at: <https://www.researchgate.net/publication/372724328>
- Salter, M. & Whitten, T. (2022). A comparative content analysis of pre-internet and contemporary child sexual abuse material', *Deviant Behavior*, 43, 1120 – 1134. <https://dx.doi.org/10.1080/01639625.2021.1967707>
- Salter, M. & Woodlock, D. (2022). The antiepistemology of organized abuse: Ignorance, exploitation, inaction, *British Journal of Criminology*, 63, 221 - 237, <https://dx.doi.org/10.1093/bjc/azac007>
- Seto, M., Buckman, C., Dwyer, R., & Quayle, E. (2018). *Production and active trading of child sexual exploitation images depicting identified victims: NCMEC/Thorn research report*. Alexandria, VA: NCMEC
- Slane, A., Martin, J., Rymer, J., Eke, A., Sinclair, R., Charles, G., & Quayle, E. (2018). Professionals' perspectives on viewing child sexual abuse images to improve response to victims. *Canadian Review of Sociology*, 579-596. <https://doi.org/10.1111/cars.12223>
- Steel, C., Newman, E., O'Rourke, S., & Quayle, E. (2020). An integrative review of historical technology and countermeasure usage trends in online child sexual exploitation material offenders. *Forensic Science International: Digital Investigation*, 33, [300971]. <https://doi.org/10.1016/j.fsidi.2020.300971>
- Steel, C., Newman, E., O'Rourke, S., & Quayle, E. (2022). Public perceptions of child pornography and child pornography consumers. *Archives of sexual behavior*, 51(2), 1173-1185. <https://doi.org/10.1007/s10508-021-02196-1>

- Steel, C. M. S., Newman, E., O'Rourke, S., & Quayle, E. (2022). Improving child sexual exploitation material investigations: Recommendations based on a review of recent research findings. *The Police Journal: Theory, Practice and Principles*. <https://doi/10.1177/0032258X221142525>
- Steel, C., Newman, E., O'Rourke, S., & Quayle, E. (2022). Technical behaviours of child sexual exploitation material offenders. *Journal of Digital Forensics, Security and Law*, 17. <https://commons.erau.edu/jdfsl/vol17/iss1/2/>
- Steel, C., Newman, E., O'Rourke, S., & Quayle, E. (2023). Lawless space theory for online child sexual exploitation material offending. *Aggression and Violent Behavior*, 68, [101809]. <https://doi.org/10.1016/j.avb.2022.101809>
- Wolak, J., Finkelhor, D., Mitchell, K. J. & ybarra, M. (2008). Online “predators” and their victims: Myths, realities, and implications for prevention and treatment. *American Psychologist*, 63(2), 111-128. <https://doi: 10.1037/0003-066X.63.2.111>
- Woodlock, D., Salter, M., Dragiewicz, M., & Harris, B. (2023). “Living in the Darkness”: Technology-facilitated coercive control, disenfranchised grief, and institutional betrayal, *Violence Against Women*, 29, pp. 987 - 1004, <http://dx.doi.org/10.1177/10778012221114920>
- Ybarra, M.L., Mitchell, K.J., Hamburger, M., Diener-West, M., & Leaf, P.J. (2011). X-rated material and perpetration of sexually aggressive behavior among children and adolescents: is there a link? *Aggressive Behaviour*, 37(1):1-18. doi: 10.1002/ab.20367. PMID: 21046607.

4) Question 4: What approach do you think we should take to the level of detail in the Code? What role could non-binding guidance play in supplementing the Code?

It is our view that a high level of detail will be required in this code due to history of a lack of interest in online technology companies in addressing concerns regarding safety and an increase in the impact of online harms (Salter & Hanson, 2021; Salter & Sokolov, 2023). By developing a code containing a high level of detail specific to obligations of VSPS providers this can act as a roadmap for providers in supporting them to take measures to meet their obligations. Loss of advertising revenue and public pressure appears to have been the only route through which there has been some change in this area, therefore it is clear that the code will need to be detailed in its expectations regarding regulation.

Non-binding guidance

Currently we are unclear as to what role this could play and reference the current Irish guidelines that state that platforms should have age restrictions of over 16, however this does not appear to be adhered to or effectively enforced.

5) Question 5: What do you think would be the most effective structure for the Code? What are the most important factors we should consider when we decide how to structure the Code?

It is our view that the most effective structure for the code is one that is split into sections as outlined on page 10 of the accompanying document. The factors that require careful consideration are, ensuring clarity and transparency that the detail contained in the code is accessible, particularly with regards to the use of language. It is our view that the language used should be simplified to ensure that the code cannot be easily misinterpreted. This may require the provision of supplementary documentation if necessary.

6) Question 6: How should we design the Code to minimise the potential for conflict and maximise the potential for synergies in how platforms comply with it and the DSA?

It is essential to ensure that the obligations of the Code are clear. The consequences of non-compliance should be outlined plainly in language that emphasises the rights of individuals to access safe online spaces. In order to minimise conflict the design of the code could utilise the language of the Online Safety and Media Regulation Act 2022 by defining obligations under the headings of “standards”, “practices” and “measures”. By ensuring that the code is clear in its expectations this will aid communication with platforms. Furthermore, feedback both through formal and informal mechanisms should be sought on how platforms are managing their obligations. It is important to ensure that any mechanism put in place to receive this feedback does not allow for the domination of one voice over another.

The above sentiments are further echoed in both the Irish Digital Services Bill 2023 and the EU Regulations under the Digital Services Act 2022. These legislative frameworks outline the obligation that VSPS’s must abide by, with the objective of providing a “safe, predictable and trustworthy online environment”. They should be seen as a collaborative resource to further enhance and strengthen the code, rather than one that has the potential for conflict. It is our recommendation that the code is aligned with these frameworks, to ensure maximum synergy, thus promoting the safety needs of individuals accessing VSPS’s, while also holding these platforms to account.

Question 7: To what extent, if at all, should the Code require VSPS providers to take measures to address content connected to video content?

Where a VSPS has any link from their platform to any other platform/ content they should have responsibility to ensure that it does not pose a risk of violent/ criminal/ sexual harm to the service user. For example, if the VSPS is a children’s platform, any link that is attached to that platform should only be to content that is not harmful and the VSPS should be responsible for ensuring this as it essentially has provided a pathway to this.

7) Question 8: How should we ask VSPS providers to introduce a feature that allows users to declare when videos contain advertising or other type of commercial communications? Should the Code include specific requirements about the form in which the declaration should take? What current examples are there that you regard as best practice?

Content creators should have to declare this before uploading which should trigger a mechanism which prompts a banner clearly indicating that it is a commercial endeavour. The onus is both on the VSPS in terms of moderating content uploaded as well as the content creator and paying organisation to ensure this occurs. VSPS providers should remove content that does not indicate this clearly if it becomes aware of any issues pending review. The experience of INHOPE, the global network of 52 hotlines, should inform the development of mechanisms for taking down material that is harmful to children (<https://www.inhope.org/EN>).

8) Question 9: How should we ask VSPS providers to introduce and design a flagging mechanism in the Code? How can we ensure that VSPS providers introduce the mechanism in a user-friendly and transparent way? How should we ask VSP Providers to report the decisions they've made on content after it has been flagged? To what extent should we align the Code with similar provisions on flagging in the DSA?

A flagging mechanisms should exist, where if a user is exposed to or engages in content that is inappropriate or causes concern, they can click to report this. Subsequent to any video being reported, access to this should be automatically stopped until a moderator has reviewed the concern and content, ensuring that it does not present a risk of harm.

VSPS providers should be obliged to engage with an auditing process, completed by an Independent Moderator; who is connected with the Commission. This will ensure that the VSPS providers are transparent in their rationale as to what determines an outcome, subsequent to a report on a video being made. The findings of the audit should then be published, with recommendations made. This process should be underpinned by statutory powers.

It is our view that VSPS providers may need to employ or seek the input of a professional with child protection expertise. All child protection policies should be provided to Tusla, Child and Family Agency and reviewed independently

9) Question 10: What requirements should the Code include about age verification and age assurance? What sort of content should be shown by default to users who are logged out or in private browsing mode and whose age cannot be verified or assured? What evidence is there about the effectiveness of age estimation techniques? What current practices do you regard as best practice? Where accounts are not age verified should default privacy settings be used, should content default to universal content and should contact by others be more limited?

The idea of a digital passport/identity appears promising and it is clear that there is a need for regulation and authentication when it comes to age verification and assurance. There needs to be safeguards to ensure age verification prior to allowing access to content. Best practice is to have some mechanism, where in alignment with GDPR, a parent has to provide proof of identification and provide consent for what age group or category is appropriate for their child. A good example is that used by online banking systems e.g. Revolut. A child is not permitted

to have an account without a parent account to which the child's account is linked; the link to the parent account allows the child to access their account. In this way, only individuals known to the parent can transfer money to the child. The protection needs to be twofold: a parent needs to give consent to ensure their child only has access to an appropriate category of content and the VSPS then needs to ensure that the child can only access to that category of content. Careful consideration of GDPR needs to occur when considering such an approach, to ensure that the data of children is not stored or used by companies for purposes other than ensuring a safe online environment. Penalties for doing so should be impactful.

In private browsing, there should be a default that you receive no age inappropriate content.

10) Question 11: What requirements should the Code have in relation to content rating? What do you consider to be current best practice? What experiences have you had using content rating systems on platforms and do you think they have been effective? What steps could we ask VSPS to take to ensure content is rated accurately by users?

We are of the view that the current model used by for example IFCO is a good example of appropriate age rating. While users may recommend an age rating the responsibility to ensure that this is in fact appropriate needs to lie with the VSPS who provides the mechanism to share content online.

11) Question 12: What requirements should the Code have in relation to parental control features? How can we ensure that VSPS providers introduce the mechanism in a user-friendly and transparent way? Can you point to any existing example of best practice in this area? Should parental controls be 'turned-on' by default for accounts of minors or where age is not verified?

It is our view that the Code should contain an obligation for parental controls to exist. We are also of the view that this should incorporate features such as, flagging, time limits, content alerts or attempts to change passwords/create new accounts. We are of the view that parental controls should be turned on automatically.

12) Question 13: What requirements should the Code contain to ensure that VSPS provide for effective media literacy measures and tools?

VSPS should be obliged to work and provide support to agencies/organisations who seek to support the development of media literacy and to independently develop measures and tools. Corporate Social Responsibility (CSR) is an important consideration and we recommend that the United Nations guidance for corporations on participation in child abuse prevention, guides this part of the Code

https://www.ohchr.org/sites/default/files/documents/publications/guidingprinciplesbusinesshr_en.pdf.

13) Question 14: How should we ask VSPS providers to address online harms in their terms and conditions in the Code, including the harms addressed under Article 28b? How should key aspects of terms and conditions be brought to users' attention? What examples are there of best practice in relation to terms and conditions including content moderation policies and guidelines?

Content should be reviewed and categorised by moderators before being available to view. A safeguard where if a video is posted on a video sharing platform that it has to be filtered through a moderating system before being accessed by others.

The recently published (9 August 2023) Ofcom report "*Regulating Video Sharing Platforms (VSPs): What we've learnt about VSPs' user policies*" also provides valuable information about how to approach this and develop best practice models.

14) Question 15: How should we ask VSPS providers to address content moderation in the Code? Are there any current practices which you consider to be best practice? How should we address automated content detection and moderation in the Code?

We suggest that VSPS are supported through the provision of training modules to help their employees understand the impact of online harms on children. Supervision of those who undertake the role of moderators of such content is, in our view, the best way to support staff and ensure they maintain their knowledge and skills in implementing a code of practice. VSPS need to be aware of vicarious trauma/secondary traumatisation/burnout and how this can impact on individuals undertaking these roles. Review mechanisms are needed to ensure that moderation is effective both in removing harmful content and reinstating content that was reported erroneously.

15) Question 16: What requirements should the Code include about procedures for complaint-handling and resolution, including out-of-court redress or alternative-dispute resolution processes? To what extent should these requirements align with similar requirements in the DSA? What current practices could be regarded as best practice? How frequently should VSPS providers be obliged to report to the Commission on their complaint handling systems and what should those reports contain? Should there be a maximum time-period for VSPS providers to handle user complaints and if so, what should that period be?

Please see response to question 10. We recommend that VSPS should report to the Commission twice yearly and that the reports should detail complaints and responses as per the classification system recommended above.

16) Question 17: What approach do you think the Code should take to ensuring that the safety measures we ask VSPS providers to take are accessible to people with disabilities?

Language and terminology used by VSPS providers should be simplified. Technological supports should be available to assist those who may require alternative means of communication, for example the use of imagery or voice overs.

17) Question 18: What approach do you think the Code should take to risk assessments and safety by design? Are there any examples you can point us towards which you consider to be best practice?

Safety by design is the principle that safety is inherent when building the product to eliminate risks. VSPS providers should be asked to provide clear information regarding how this has been part of the design of their platform. A good example of safety by design can be found in the healthcare space where safety by design and risk assessment is a core aspect of practice, subject to regulation, before online products are utilised in this space.

18) Question 19: How do you think that cooperation with other regulators and bodies can help us to implement the Code for VSPS?

We are of the view that in order to foster cooperation, ongoing dialogue is essential. The use of workshops and training will also be an important mechanism to foster cooperation. In order to bring about effective implementation of the Code, transparency regarding compliance is necessary. This could take the form of publication of how VSPS providers have complied with the code written in a clear, simple and accessible way. Furthermore, the alignment of the code with existing regulation will support cooperation.

19) Question 20: What approach do you think we should take in the Code to address feeds which cause harm because of the aggregate impact of the content they provide access to? Are there current practices which you consider to be best practice in this regard?

We are concerned about the influence of the aggregate impact on children, particularly in the area of child sexual abuse and mental health. It is our view that mechanisms to address this need to be developed. We suggest exploring the possibility of creating a flagging system to detect this content where children have accessed inappropriate content, and manipulating the algorithm to enable the VSPS to interrupt the flow of aggregate content when such content presents risk harm. An example of this is when a message is flagged ‘do you want to proceed’, leading to an alert being sent to a parent’s phone and parental consent would be required if the child wishes to proceed. A mechanism whereby parents would be sent summaries of websites accessed, such is currently used by service providers to inform users of average screen time in a day or week could be used.

20) Question 21: Do you have any views on how requirements for commercial content arranged by a VSPS provider itself should be reflected in the Code?

VSPS providers should adhere to international best practice, such as the UN guidance (<https://www.ohchr.org/sites/default/files/documents/publications/guidingprinciplesbusinessh>)

[r_en.pdf](#)) and this requirement should be reflected in the code. Transparency regarding commercial content is required.

21) Question 22: What compliance monitoring and reporting arrangements should we include in the Code?

External audit of VSPS is required to ensure monitoring and that reporting arrangements are adhered to.

22) Question 23: Should the Code have a transition period or transition periods for specific issues? Which areas touched on in this Call for Inputs may VSPS providers require time to transition the most? What time frame would be reasonable for a transition period?

It is our view that Very Large Online Platforms (VLOPs) should be tasked with implementing the Code immediately. There have been many opportunities for VSPS to address the known impact and safety issues prolific in this space, yet this has not been actively pursued. A short transition period for smaller providers could be in place, however, this should be minimal. We suggest the implementation of a review following the transition to enable VSPS and other interested parties to provide feedback on the implementation of the Code.

**Submission to
Coimisiún na Meán's Call For Inputs:
Developing Ireland's First Binding
Online Safety Code for Video-Sharing
Platform Services**

August 2023



Women's  Aid



Contents

About Women's Aid..... 4

Introduction..... 4

Answers to consultation questions 7

 Question 1: What do you think our main priorities and objectives should be in the first binding Online Safety Code for VSPS? What are the main online harms you would like to see it address and why? 7

 Question 2: What types of online harms do you think should attract the most stringent risk mitigation measures by VSPS? How could we evaluate the impact of different types of harms e.g. severity, speed at which harm may be caused? Is there a way of classifying harmful content that you consider it would be useful for us to use? 10

 Question 3: Do you have reports, academic studies or other relevant independent research that would support your views? If you do, please share them with us with links to relevant reports, studies or research..... 11

 Question 4: What approach do you think we should take to the level of detail in the Code? What role could non-binding guidance play in supplementing the Code? 12

 Question 5: What do you think would be the most effective structure for the Code? What are the most important factors we should consider when we decide how to structure the Code? 13

 Question 6: How should we design the Code to minimise the potential for conflict and maximise the potential for synergies in how platforms comply with it and the DSA? 14

 Question 7: To what extent, if at all, should the Code require VSPS providers to take measures to address content connected to video content? 15

 Question 8: How should we ask VSPS providers to introduce a feature that allows users to declare when videos contain advertising or other type of commercial communications? Should the Code include specific requirements about the form in which the declaration should take? What current examples are there that you regard as best practice? 16

 Question 9: How should we ask VSPS providers to introduce and design a flagging mechanism in the Code? How can we ensure that VSPS providers introduce the mechanism in a user-friendly and transparent way? How



should we ask VSP Providers to report the decisions they've made on content after it has been flagged? To what extent should we align the Code with similar provisions on flagging in the DSA? 16

Question 10: What requirements should the Code include about age verification and age assurance? What sort of content should be shown by default to users who are logged out or in private browsing mode and whose age cannot be verified or assured? What evidence is there about the effectiveness of age estimation techniques? What current practices do you regard as best practice? Where accounts are not age verified should default privacy settings be used, should content default to universal content and should contact by others be more limited? 19

Question 11: What requirements should the Code have in relation to content rating? What do you consider to be current best practice? What experiences have you had using content rating systems on platforms and do you think they have been effective? What steps could we ask VSPS to take to ensure content is rated accurately by users? 20

Question 12: What requirements should the Code have in relation to parental control features? How can we ensure that VSPS providers introduce the mechanism in a user-friendly and transparent way? Can you point to any existing example of best practice in this area? Should parental controls be 'turned-on' by default for accounts of minors or where age is not verified? .. 21

Question 13: What requirements should the Code contain to ensure that VSPS provide for effective media literacy measures and tools? 21

Question 14: How should we ask VSPS providers to address online harms in their terms and conditions in the Code, including the harms addressed under Article 28b? How should key aspects of terms and conditions be brought to users' attention? What examples are there of best practice in relation to terms and conditions including content moderation policies and guidelines? 22

Question 15: How should we ask VSPS providers to address content moderation in the Code? Are there any current practices which you consider to be best practice? How should we address automated content detection and moderation in the Code? 24

Question 16: What requirements should the Code include about procedures for complaint-handling and resolution, including out-of-court redress or alternative-dispute resolution processes? To what extent should these requirements align with similar requirements in the DSA? What current



practices could be regarded as best practice? How frequently should VSPS providers be obliged to report to the Commission on their complaint handling systems and what should those reports contain? Should there be a maximum time-period for VSPS providers to handle user complaints and if so, what should that period be? 26

Question 17: What approach do you think the Code should take to ensuring that the safety measures we ask VSPS providers to take are accessible to people with disabilities? 27

Question 18: What approach do you think the Code should take to risk assessments and safety by design? Are there any examples you can point us towards which you consider to be best practice? 28

Question 19: How do you think that cooperation with other regulators and bodies can help us to implement the Code for VSPS? 30

Question 20: What approach do you think we should take in the Code to address feeds which cause harm because of the aggregate impact of the content they provide access to? Are there current practices which you consider to be best practice in this regard? 30

Question 21: Do you have any views on how requirements for commercial content arranged by a VSPS provider itself should be reflected in the Code? 31

Question 22: What compliance monitoring and reporting arrangements should we include in the Code? 31

Question 23: Should the Code have a transition period or transition periods for specific issues? Which areas touched on in this Call for Inputs may VSPS providers require time to transition the most? What time frame would be reasonable for a transition period? 32

Conclusions..... 32

About Women's Aid

Women's Aid is a national, feminist organisation working to prevent and address the impact of domestic violence and abuse (henceforth referred to as DVA) including coercive control, in Ireland since 1974. We do this by advocating, influencing, training, and campaigning for effective responses to reduce the scale and impact of DVA on women and children in Ireland and providing high quality, specialised, integrated, support services. More information on Women's Aid is available on our website www.womensaid.ie

Introduction

Women's Aid welcomes the establishment of an Online Safety Commissioner to oversee the new regulatory framework for online safety and is pleased to provide a submission to Coimisiún na Meán on the Call for Inputs: Developing Ireland's First Binding Online Safety Code for Video-Sharing Platform Services.

Cyber-stalking and Imaged Based Sexual Abuse (IBSA) have been a great concern for Women's Aid over a number of years. In the context of DVA, IBSA is used as a tactic to control, humiliate and harass a partner or ex-partner.

Many women have told us that their partner or ex-partner has taken and/or published sexually explicit images of the woman without her consent, damaging her reputation, self-esteem and possibly work opportunities and relationships. The perpetrators use these images to threaten, blackmail, and humiliate the woman, especially if she has

indicated her desire to end the relationship or has already done so. In other cases, he uses the images for his financial gain without the woman's knowledge or consent by uploading them onto commercial websites. In some cases, her contact details (including phone, address, and social profiles) are also published, for example on escort websites.

Regardless of the motive, this type of abuse has huge negative impact on the woman and may cause immense and irreversible harm. The more identifiable a woman is, the more devastating is the impact on her of having these images published/distributed.

Young women are more likely to suffer cyber-abuse and specifically image based sexual abuse:

- 1 in 5 young women experience intimate relationship abuse in Ireland.
 - Nearly half (49%) of whom experience online abuse by their partners and ex-partners.
 - Of these, 20% had images or videos taken of them without their permission with 15% having been threatened with sharing sexually explicit intimate photos and or videos and 17% having actually had sexually explicit or intimate videos or images shared without their consent.¹
- Hotline.ie reports that in the period September 2021 to September end 2022 they received 773 reports of intimate image abuse.

¹ Women's Aid, 2020 One in Five young women suffer intimate relationship abuse in Ireland. Available here: <https://www.womensaid.ie/app/uploads/2023/04/One-in-Five-Young-Women-Report-2020.pdf>

- of which 525 were actionable.
- For 90% of the reports, they were successful in having the images removed.
- 83% of the people reporting intimate image abuse were women and the great majority was under 35 years old.²

Given the above, the most pressing issue for us in relation to Video-Sharing Platform Services (henceforth VSPS) is the **non-consensual sharing of intimate images/videos and the comments posted about them**, which are often degrading, sometimes violent, and can compound the negative impact on women and girl's mental health and wellbeing.

We therefore welcome the drafting of binding codes for VSPS. In particular we would like to comment on the prevention of uploading and sharing intimate videos without consent and the response of platform services when reports are made to them by victims/survivors.

An important part of this response should be fast and free take downs. The Harassment, Harmful Communications and Related Offences Act 2020 created much needed offences in relation to image-based sexual abuse. However criminal prosecutions take time and, for a variety of reasons, do not always go ahead.³ In the meantime, the images are available and can be shared and re-posted numerous times. The more

² Hotline.ie. (2022). Hotline.ie 2021 Annual Report. Dublin, Ireland

³ McGlynn C, e al. 2019, Shattering Lives and Myths: A Report on Image-Based Sexual Abuse

https://www.researchgate.net/publication/339352950_Shattering_lives_and_myths_A_report_on_image-based_sexual_abuse



IBSA material is allowed to go viral, the more difficult it is to eliminate it from the Internet and the more harm that is done.

For the majority of victims, swift removal of intimate videos/images shared without consent is a priority and more important than prosecution. For example, Hotline.ie reports that “only 1 in 7 reporters indicated they wished to have the matter referred to An Garda Síochána for law enforcement investigations. The vast majority opted for content removal only”.⁴

Getting images/videos removed from the internet can be difficult, costly and time-consuming, and it should **not** be the responsibility of the victim/survivor.

We have provided answers to the questions which are most relevant to our remit and concerns as detailed above.

Answers to consultation questions

Question 1: What do you think our main priorities and objectives should be in the first binding Online Safety Code for VSPS? What are the main online harms you would like to see it address and why?

As outlined above our main concern regarding VSPS is the non-consensual sharing of intimate images/videos, including altered/fake ones, which are becoming more and more common.

⁴ Hotline.ie (2022) op. cit

Once an intimate video is uploaded, it can go viral and be shared multiple times. It then becomes nearly impossible to delete all occurrences, and even if the video is deleted from the original site, it can reappear on others endlessly, which is extremely harmful to the victims.

Intimate images are shared without consent on a variety of platforms, and many are shared on video sharing platforms. According to Hotline.ie, 51% of this imagery reported to them was shared on video streaming services and 23% of image hosting services.⁵

Priorities for this code therefore should be to:

- prevent the uploading or sharing of intimate videos/content unless **consent has been verified prior** to the uploading/sharing. This means that anonymous account should not be able to upload or share this content and that users will have to confirm they are sharing with consent.

For example, the ERAW Violence against Women and Girls Code of Practice for suggests that to mitigate harm of IBSA in platforms with user generated or uploaded pornography, “services should require user verification before uploads and require users to confirm they have consent from everyone depicted in the content to upload. This should be accompanied with messaging that informs them it is a criminal offence to upload material without the consent of those depicted, including content in violation of

⁵ Hotline,ie op. cit.

copyright and that the platform will take action against users for doing this.”⁶

- Address the impact of deep-fake pornography by including it in any such requirements.
- Require clear, fast take down procedures for platforms, provided at no cost to the user, with penalties for not doing so within strict timeframes.
- Require platforms to also have to delete links or comments linked to intimate videos posted without consent.
- Require platforms to raise awareness about the harm and unacceptability of sharing intimate images/videos without consent.
- Address the way multiple forms of discrimination intersect and intensify the negative impact of abuse in the experiences of marginalized individual and groups.

Note that a huge percentage of images is shared without consent to adult pornographic sites,⁷ it is therefore essential that they are included

⁶ EAW Violence Against Women and Girls (VAWG) Code of Practice, page 16; <https://drive.google.com/file/d/1cMIginaMEN2kULCL2eftH2B7oGVK9FZh/view>

⁷ In the UK, the Revenge Porn Hot-line estimates that “Private sexual content is frequently shared on adult content sites, in around 40% of cases where content is shared” Ward, Revenge Porn Helpline Report 2022, <https://revengepornhelpline.org.uk/resources/helpline-research-and-reports/>

Similarly the Australian e-safety commissioner reports the majority of IBSA material was posted on exposé or pornography sites Australian government, ACMA and eSafety Annual reports 2021-2022 page 183 <https://www.esafety.gov.au/sites/default/files/2022-10/ACMA%20and%20eSafety%20annual%20report%202021-22.pdf>

in the codes. Moreover, given the increase use of deep-fakes in image based sexual abuse, this should also be specifically included.⁸

The definition of intimate image should correspond to the Harassment, Harmful Communications and Related Offences Act 2020.

Other harmful content of concern includes misogynistic videos (for example relating to incel) and channels where perpetrators of abuse seek suggestions and guidance to help them abuse, which should also be included in the code for action.

Question 2: What types of online harms do you think should attract the most stringent risk mitigation measures by VSPS? How could we evaluate the impact of different types of harms e.g. severity, speed at which harm may be caused? Is there a way of classifying harmful content that you consider it would be useful for us to use?

Online Violence against women has severe impacts on victims/survivors, affecting their mental health, physical safety in the real world, reputation, relationships, and employment and their ability/willingness to maintain an online presence.⁹

Image based sexual abuse (IBSA) content, should be a priority, as it is extremely harmful, as confirmed by numerous studies and by our own

⁸ The State of Deepfakes: Landscape, Threats, and Impact, Henry Ajder, Giorgio Patrini, Francesco Cavalli, and Laurence Cullen, September 2019.

⁹ See for example, the Guardian, 'There's no end and no escape. You feel so, so exposed': life as a victim of revenge porn, 22 September 2019, <https://www.theguardian.com/lifeandstyle/2019/sep/22/theres-no-end-and-no-escape-you-feel-so-so-exposed-life-as-a-victim-of-revenge-porn>

experience supporting victim/survivors. Women are disproportionately impacted by online abuse and IBSA in particular. Marginalised women even more so.

In relations to Imaged based abuse and the sharing of intimate content without consent, it is important to note that the more the victim is identifiable the worse the harm. So cases where personal information is also shared with the image/video (for example name, address, social media profiles) or where the person is easily identifiable (for example clearly visible face), this should be prioritised.

Where the content shared is a recording of rape/sexual abuse and/or involves children, this would be an absolute priority.

Question 3: Do you have reports, academic studies or other relevant independent research that would support your views? If you do, please share them with us with links to relevant reports, studies or research.

- McGlynn C, e al. 2019, "Shattering Lives and Myths: A Report on Image-Based Sexual Abuse."

https://www.researchgate.net/publication/339352950_Shattering_lives_and_myths_A_report_on_image-based_sexual_abuse

- Plan International, 2020, "Free to be online? Girls' and young women's experiences of online harassment." <https://plan-international.org/publications/free-to-be-online/>

- Glitch, UK (2023) “The Digital Misogynoir Report: Ending the dehumanising of Black women on social media.”
www.glitchcharity.co.uk/research

While not specific to VSPS, the publications below offer very useful considerations for designing codes that address VAW online:

- ERAW Violence Against Women and Girls (VAWG) Code of Practice
<https://www.endviolenceagainstwomen.org.uk/wp-content/uploads/2022/05/VAWG-Code-of-Practice-16.05.22-Final.pdf>
- End Cyber Abuse, Orbits A field guide to advance intersectional, survivor-centred, and trauma-informed interventions to tech abuse (technology-facilitated gender-based violence)
<https://endcyberabuse.org/orbits/>

Question 4: What approach do you think we should take to the level of detail in the Code? What role could non-binding guidance play in supplementing the Code?

Women's Aid believes the code should be quite detailed and prescriptive and therefore we would not recommend approach 2 (a very High-level code).

Our experience, with women contacting platforms to have material taken down, is that it can be frustrating and traumatizing, with women not knowing what to do, who to contact /reporting channels, not getting responses, not knowing timeframes for actions or their rights. It seems

that platforms do not always enforce even their own regulations, especially if harmful content draws a lot of views.¹⁰ Therefore codes need to be enforceable.

We believe the code should be very clear and prescriptive in regard to the responsibilities of VSPS both in terms of prevention and in terms of action when reports are made.

The code should include a commitment to work with hotline.ie and equivalent services in other jurisdictions in relation to removal of CSA and IBSA content.

Non-binding guidance platforms are welcome to help ensure consistency and clarity but there needs to be enforcement of the code, binding rules are therefore more important.

We welcome Coimisiún na Meán's plan to introduce an accompanying guide, and recommend that the guide is inclusive and accessible to all users, including young people.

Question 5: What do you think would be the most effective structure for the Code? What are the most important factors we should consider when we decide how to structure the Code?

As the range of content the code will address is quite varied and the level of harm different, Women's Aid believes that it would be useful for the structure of the code to have a separate section for each main

¹⁰https://www.nbcnews.com/tech/internet/deepfakes-twitter-tiktok-stars-rcna87295?mc_cid=adf5fa2110&mc_eid=1fd5b6746d

category of content it addresses. This would make the code specific and clear regarding how each harm is addressed.

Image based sexual abuse would have to be one of the major separate sections.

Within each major section there could be similar subsections addressing the relevant measures (Content Policies / T&Cs; Risk Assessments; Content Moderation and Complaints; Online Safety Features; Service Design Measures; Compliance Measures.)

Question 6: How should we design the Code to minimise the potential for conflict and maximise the potential for synergies in how platforms comply with it and the DSA?

Women's Aid agrees the code should maximise synergies with the DSA. While we do not have firm suggestions regarding design the code should be designed with the objective to:

- Require commitments (and ensure mechanisms to evaluate) co-operation between platforms to minimize the burn out on a victim/survivor having to deal with multiple platforms relating to a single or connected experience of online harm.
- Require all platforms to have - or sign up to - a meaningful commitment to recognize specific gendered violence and harm that can be affected and perpetuated against women and girls on their platforms. This should include acknowledgement and recognition of intersectional factors which exacerbate harms to

women and girls from minoritized backgrounds and circumstances.

Question 7: To what extent, if at all, should the Code require VSPS providers to take measures to address content connected to video content?

As mentioned, there could be content linked to the intimate images shared without consent which could be extremely harmful:

- content that may identify or locate the person or content that falsely suggests the person provides sexual services. For example, women report to us their partners post videos of them on escorts sites, without the woman's consent or knowledge and include their phone number, social media profiles or address. This should also include incidents of 'Doxing' (sharing of personal information about an individual online with a malicious intention) which can include, for example, sharing a video of someone's home and threatening to - or inciting others to - go to their home and harass or do them harm.
- Derogatory, offensive, threatening and abusive comments often features on the sites where intimate videos are posted without consent and increase the victim's trauma.

Women's Aid recommends that the code should provide that:

- Where there is a request for a video to be taken down, all related content and links should also be deleted.

- in any case abusive, misogynistic and violent comments should not be allowed and platforms should be required to develop policies recognizing gendered violence and abuse; setting out both their commitments to eliminating this - and tangible actions to address this in the round on their platforms.

Question 8: How should we ask VSPS providers to introduce a feature that allows users to declare when videos contain advertising or other type of commercial communications? Should the Code include specific requirements about the form in which the declaration should take? What current examples are there that you regard as best practice?

Not in our remit.

Question 9: How should we ask VSPS providers to introduce and design a flagging mechanism in the Code? How can we ensure that VSPS providers introduce the mechanism in a user-friendly and transparent way? How should we ask VSP Providers to report the decisions they've made on content after it has been flagged? To what extent should we align the Code with similar provisions on flagging in the DSA?

Women's Aid believes that flagging/ reporting mechanisms need to be **visible**, transparent, accessible and free for any users.

As many of the VSPS based in Ireland have an international/global presence, it is vital that these mechanisms are accessible in the local

language/s of the user. It is not good enough for them to be only in English. They should also be designed with the needs of children, young people and people with additional needs and/or disabilities in mind.

Information on reporting mechanisms, detailing what can be expected by the VSPS after a report is made and within which time-frames need to be provided in accessible formats including plain (local) language/s, and need to be easy to locate on the website/platform.

Once a user flags IBSA content, the user should also be shown a message acknowledging the report and summarising what would happen next. The message should also include information on relevant and local (to the country) supports and on the Online Safety Commissioner/equivalent. This should be done considering the safety of the reporter, for example this information should not be automatically retained in the browser or the account of the user.

There should also be an option for offline reporting (phone line) to ensure survivors whose access to the Internet is controlled or monitored by the abuse can report image based sexual abuse safely.

Moreover, there should be options for users with disabilities, for example there should be the possibility to make voice-activated reporting mechanisms for users who may have visual impairments or literacy issues.

Users should be informed of the decision made and reasons for it regarding the flagged content. The way to receive this information should be chosen by the user to maintain their safety.

Women's Aid agrees that it seems a good idea to integrate the flagging mechanism under the DSA and the Code, as this would be a more user-friendly option than having two different mechanisms.

The DSA (Article 16) will require platforms to put in place a notification mechanism for illegal content and require them to process the notifications in a timely, diligent, non-arbitrary and objective manner. This should be integrated into the Code being developed. It is important to make the process for flagging content as straightforward and easy to understand for children and young people as possible. Children in particular may find some of the rules set out in community guidelines confusing or struggle to distinguish between what is illegal and what is legal but prohibited by a service. Requiring users to determine whether they are flagging content under the DSA or the Code would place a significant burden on the user and could act as a deterrent to children and young people flagging illegal and harmful online content.¹¹

¹¹ Children's Rights Alliance Submission on Online Safety Code.

Question 10: What requirements should the Code include about age verification and age assurance? What sort of content should be shown by default to users who are logged out or in private browsing mode and whose age cannot be verified or assured? What evidence is there about the effectiveness of age estimation techniques? What current practices do you regard as best practice? Where accounts are not age verified should default privacy settings be used, should content default to universal content and should contact by others be more limited?

Women's Aid agrees on user age verification for certain content. Pornography is widely available to children and young people and shapes their understanding of sex and relationships. It harms both girls and boys, by influencing expectations, normalising disrespectful sexual behavior and promoting misogynistic, and often abusive and violent, models of sexual expectation.

Recent Women's Aid research found that:

- The majority of Irish people believe that pornography is too accessible to children, and that it is contributing to gender inequality and to coercion and sexual violence against women and girls.
- 73% of respondents believe that we must end children' and young people's exposure to pornography if we are to foster healthy sex and intimate relationships.

- 75% of people believe that pornography makes children and young people more vulnerable to requests for sexually explicit images and videos.¹²

Age verification is therefore an essential tool, however we do not have an opinion regarding the best technology to be used. We also recommend that access to adult content to users whose age cannot be verified should be restricted. Women's Aid also stress that age verification **alone** cannot be considered a 'panacea'/the only mechanism to protect children and young people and must be considered as one of a range of protective mechanisms.

Question 11: What requirements should the Code have in relation to content rating? What do you consider to be current best practice? What experiences have you had using content rating systems on platforms and do you think they have been effective? What steps could we ask VSPS to take to ensure content is rated accurately by users?

No comment.

¹² Women's Aid, 2022, It's time to talk about porn Irish attitudes on the links between pornography, sexual development, gender inequality and violence against women and girls. Available here:
https://www.womensaid.ie/app/uploads/2023/06/its_time_to_talk_about_porn_report_womens_aid_november_2022.pdf

Question 12: What requirements should the Code have in relation to parental control features? How can we ensure that VSPS providers introduce the mechanism in a user-friendly and transparent way? Can you point to any existing example of best practice in this area? Should parental controls be 'turned-on' by default for accounts of minors or where age is not verified?

Women's Aid believes that all online platforms should be safe for everyone. Further, we also believe that the onus of safety should be with the online platform. It should be the responsibility of platforms to ensure that the onus does not fall on users to utilize safety settings, and that their platform is a safe, respectful environment. This should be the case for children and adults alike.

We recommend that safety and privacy setting for minors should be set at maximum safety and privacy by default.

Question 13: What requirements should the Code contain to ensure that VSPS provide for effective media literacy measures and tools?

No comment.

Question 14: How should we ask VSPS providers to address online harms in their terms and conditions in the Code, including the harms addressed under Article 28b? How should key aspects of terms and conditions be brought to users' attention? What examples are there of best practice in relation to terms and conditions including content moderation policies and guidelines?

Terms and Conditions are very important to make it clear to the users what kind of online behaviour will not be tolerated. They need to be clear and simple, in local languages and accessible to all users. They should not be too long or legalistic, as users will simply not read them and users should sign up to them before being able to upload content, comment, or be an active user on the platform.

We agree that a summary in simple language would be useful and also periodical reminders, particularly if there has been any updates.

Terms and Conditions should make clear the platform commitment to combat the spread of online violence against women and girls (VAWG), and spell out in clear language that gender based violence and misogyny online will not be tolerated.

They should outline how the service will respond to VAWG:

- Including uploading or sharing of intimate images without consent.
- which steps would be taken and commitment to short time-frames for action.



Note that in order to create awareness of non-consensual sharing of intimate images as harmful content, it is important that image-based sexual abuse is specifically **named** and made visible in the T&C and it is not “hidden” in the generic category of illegal content.

Platforms that allow adult content, should make it clear that the consent of all person depicted is necessary **prior** to uploading, and there would be consequences if this requirement is not adhered to. If possible this obligation should be made a legal requirement. We do however also note, and emphasize, that where a woman or young person is subject to coercion and exploitation that consent may ‘appear to be given’ in uploading of content, but that it can be revealed that they were coerced to do so. Therefore, it is vital that platforms recognize this and respond swiftly, and without question, to any subsequent complaint **regardless of** whether there was any initial indication of ‘consent’.

The T&C should also reference the users’ privacy rights under the GDPR, including the right to be forgotten and how to request this.

Moreover, a platform service should be responsible for, and make a written commitment to ensuring that algorithms do not suggest material that is in contravention of the site’s own Terms and Conditions.

Question 15: How should we ask VSPS providers to address content moderation in the Code? Are there any current practices which you consider to be best practice? How should we address automated content detection and moderation in the Code?

AI moderation need to be carefully deployed so that it does not operate in a discriminatory way. It cannot completely replace human moderation. There needs to be clear ways for the users to contact a human moderator if they are dissatisfied with the way automated moderation dealt with content and have the automated decision reviews within strict time-frames.

Moderators need to be trained on the various forms of online violence against women and supported in dealing with what is often harrowing and disturbing content. They also need to be culturally competent for the local areas they monitor. They need to also be trained in diversity and inclusion.

There needs to be a sufficient number of moderators appropriate to the size of the platforms. For bigger platforms there could be specific Violence against Women (VAW) moderators, with more in-depth training.

Illegal content, including image-based sexual abuse, should be taken down immediately. If there is any doubt as to whether content does or does not constitute image-based sexual abuse, the code should stipulate that **the content in question will be taken down immediately pending a final decision being made**, to prevent it going viral in the meantime.



Time-frames for taking action on reports may vary depending on the issue being raised. We note E-safety in Australia responds within a **maximum** of 2 business days, often sooner, to reports about child cyberbullying, adult cyber abuse, image-based abuse or child sexual exploitation material. It seems a fair time-frame, provided that such material is taken down pending the more detailed examination of the material in question. It can be reinstated if it is found that it is 'legitimate' content.

If survivors chose to pursue criminal or civil cases against perpetrators, the platforms should provide them promptly upon request with any evidence they have in their system.

Relevant VAW specialist services should be considered trusted flaggers in relation to IBSA and other VAW online content and content flagged by them should be immediately removed while review is pending. Services should be compensated for this role. However, they should not become the **only** flaggers, and users should be able to flag content themselves as well. Specialists VAW services could also have a role in informing the Commission about new trends in harmful VAWG content.

Question 16: What requirements should the Code include about procedures for complaint-handling and resolution, including out-of-court redress or alternative-dispute resolution processes? To what extent should these requirements align with similar requirements in the DSA? What current practices could be regarded as best practice? How frequently should VSPS providers be obliged to report to the Commission on their complaint handling systems and what should those reports contain? Should there be a maximum time-period for VSPS providers to handle user complaints and if so, what should that period be?

The code should require that platforms have clear complaint procedures, with appropriate time-frames, including a maximum period.

In particular the code should include specific guidance on complaints about decisions on illegal and harmful content, especially image-based sexual abuse. Women's Aid believes that during any dispute proceedings regarding intimate images shared without consent, such images should be taken down within a fixed, short time frame while the dispute is resolved as a precaution against further sharing, while the status of the images is determined.

Acknowledgment of complaint should be within 24 hours and should specify the next steps and how long they will take. Time-frame for the resolution of the complaint may depend on the type of complaint/s and the potential harm, in any case there should be a maximum period in which a decision is made and remediation action (if any) is completed.

When users report or complain about VAWG or image-based sexual abuse content, their contact details should not be shared with the alleged perpetrator/s. Every effort should be made to protect their data and identity from any third party.

There should be an appeal process. For image-based sexual abuse and other VAW content, the appeal should be examined by a trusted service in the trusted flaggers scheme, or the Online Safety Commissioner.

VSPS services should report on complaints handling system quarterly, they must include how many complaints were made in the period by type of complaint and how they were resolved and the time-frame in which they were solved. Complaints in relation to VAWG content should be visible separately from other types. See Question 22

Question 17: What approach do you think the Code should take to ensuring that the safety measures we ask VSPS providers to take are accessible to people with disabilities?

The best approach is to design safety measures together with people with disabilities and/or relevant services from the beginning and not as an afterthought.

However, some suggestions may include (as examples):

- Using clear and inclusive language on all communications, including T&Cs.
- Providing information in multiple formats e.g. video (with captions) as well as text.

- Providing different ways of flagging/making a complaint (voice report, third party report).

Question 18: What approach do you think the Code should take to risk assessments and safety by design? Are there any examples you can point us towards which you consider to be best practice?

Women's Aid agrees that the principle of safety by design should be included in the code, and that VSPS should carry out risk assessment of new and existing features on their platforms and how they can be abused by users to perpetrate Violence against women and girls.

For example, platforms should:

- Set users setting to maximum safety by default (with possibility to change for adult users).
- Ensure algorithms do not promote hateful content, including misogynistic content.
- require that users uploading intimate images have to confirm that they have consent of all people depicted in them, and remind them of the consequences should that not be the case. This should be a requirement for each image uploaded, not a once off.
- In relation to consent: where an individual is subject to coercion and exploitation that consent may 'appear to be given' in uploading of content, but it can be revealed that they were coerced to do so. Therefore, there must be a commitment that a platform recognize this possibility and respond swiftly to any

subsequent complaint without question, regardless of whether there was any initial indication of 'consent'.

- Ensure deep fake and nudification technology cannot be used to harm women and children on their platforms.
- Give users control on how their images/video can be downloaded and shared.
- Use digital fingerprinting, to assist with removing offending materials from all platforms and flagging accounts that shared the offending materials.
- Refer users who flag IBSA content to relevant supports in their country.
- Highlight no tolerance of VAWG and IBSA in their T&C and other relevant information.
- Provide visible and easy to access in platform report and complaints mechanisms.
- Giving survivors the option to report through an independent third party reporting platform (e.g. in Ireland hotline.ie). This would allow survivors to report IBSA content uploaded in different platforms once, rather than have to contact each platform. This option needs to be visible and accessible.

Safety by design and risk assessment need to not only focus on the individual but also consider the broader social and cultural harm of not allowing VAWG online and IBSA culture go unchallenged, and what this

means for women's and girl's safety online and offline and for women's and girl's ability to freely engage with the online world.

Question 19: How do you think that cooperation with other regulators and bodies can help us to implement the Code for VSPS?

It is important that the Commission works with other regulators at EU and global level to implement the code. In particular, at EU level clarity is needed regarding who is responsible for platforms with HQs in Ireland and the role of regulators in member states and in Ireland.

If a regulator is not the appropriate one for a complaint, the regulator should pass on the complaint to the appropriate regulator (with consent of the user making the complaint) and not ask the complainant to start anew in another jurisdiction.

Question 20: What approach do you think we should take in the Code to address feeds which cause harm because of the aggregate impact of the content they provide access to? Are there current practices which you consider to be best practice in this regard?

In certain cases individual pieces of content may not seem harmful, but a number of pieces in the aggregate, on the same or different platforms, may have great negative impact.

When content is flagged, moderators should engage with the user and consider the whole pattern of abuse including on other platforms and

offline, before making decisions regarding appropriateness of content and action to block/remove it.

Platforms should be responsible to design algorithms that do not amplify harmful contents. Platforms should also collaborate with each other both with technology and coordinated responses to create a seamless response that will minimize any need for an individual to have to engage multi-laterally with different platforms in respect of the same complaint.

Question 21: Do you have any views on how requirements for commercial content arranged by a VSPS provider itself should be reflected in the Code?

This is outside our remit.

Question 22: What compliance monitoring and reporting arrangements should we include in the Code?

In relation to VAWG and image-based sexual abuse (IBSA) content, VSPS should be required to monitor and report quarterly to the Commission on:

- Preventative measures taken to limit VAWG online and in particular to prevent the spreading of IBSA content, including risk assessment carried out.
- How many trained moderators they have available to monitor these issues specifically.

- Number of IBSA/ misogynistic videos flagged, outcomes and time-frames.
- Number of complaints received, outcomes and time-frames.
- Number of videos promoting VAWG removed.
- Number of videos with IBSA content removed.
- Number of accounts closed or blocked.
- Data should include details on race, sex/gender, gender identity and other protected characteristics of depicted victims and information on whether content was flagged automatically, by moderator, by targeted individual or third party.

Moreover, VSPS should commit to release non identifying data to bona fide researchers.

Question 23: Should the Code have a transition period or transition periods for specific issues? Which areas touched on in this Call for Inputs may VSPS providers require time to transition the most? What time frame would be reasonable for a transition period?

No comment.

Conclusions

While this code addressed VSPS specifically, the issues highlighted in this submission are relevant for other online services and social media as well. Since digital abuse is not compartmentalised and can be carried

out through different platforms there should be synergy and complementarity between codes for different types of platforms and services when these are developed in the future.

Women's Aid believes we cannot rely on platforms only to enforce codes only and that the Online Safety Commissioner should have a strong monitoring and enforcing role.

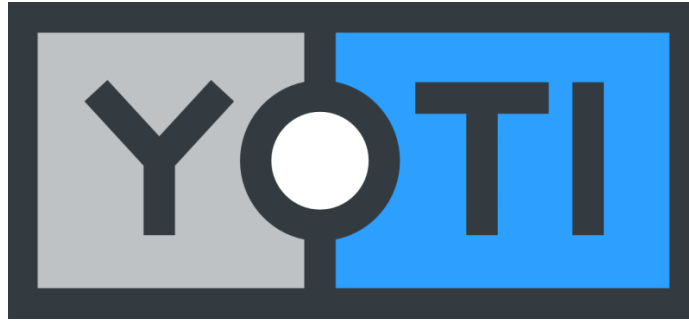
While hotline.ie has been very successful in getting content removed in the majority of cases, they only have “soft power” to do so and lack the ability to issue binding take down orders. They are also limited to residents of Ireland and public websites/platforms therefore many of the platforms and websites implicated in abuse are not currently covered by this service. EU residents wishing to make a report against a platform headquartered in Ireland cannot avail of hotline.ie and it is not clear who would be able to assist them.

Women's Aid recommends that the role of the Online Safety Commissioner is expanded to include responding to individual complaints of image-based abuse and other harmful content and facilitating their removal, at least in cases outside the remit or power of hotline.ie. Failing that, that the Online Safety Commissioner would at least have an appeal role in relation to take-down requests, as recommended in the Law Reform Commission report.¹³

¹³ Law Reform Commission, 2016, Final Report on Harmful Communications and Digital Safety, page 143, Paragraph 3.77, https://www.lawreform.ie/_fileupload/Reports/Full%20Colour%20Cover%20Report%20on%20Harmful%20Communications%20and%20Digital%20Safety.pdf



Women's Aid are grateful for the opportunity to submit on this very important piece of work and are available to discuss any aspect of our submission with Coimisiún na Meán on request.



Yoti response to the Coimisiún na Meán consultation on Online Safety Code

4 September 2023

Organisation name:

Yoti

Respondent full name:

Julie Dawson, Chief Policy & Regulatory Officer
Florian Chevoppe-Verdier, Public Policy Associate

Email address:

[REDACTED]

Contact phone number:

[REDACTED]



Contents

What requirements should the Code include about age verification and age assurance? What current practices do you regard as best practice?	3
What evidence is there about the effectiveness of age estimation techniques?	11



What requirements should the Code include about age verification and age assurance? What current practices do you regard as best practice?

We are thankful to the Media Commission for the opportunity to contribute to the development and implementation of the upcoming Online Safety Codes. As a global provider of age assurance and with experience of providing over 600 million age checks, Yoti is committed to supporting regulators and policymakers to ensure that they can build robust and inclusive online safety regimes. In response to this question, we would like to make a number of recommendations to the Commission which are listed below, and will include more information about our technology and products, and will remain available for any follow-up questions or meetings.

We will elaborate below on the following proposed requirements that the Code could include:

- Keeping with the highest possible levels of data protection & minimisation
- Choice and inclusion of age assurance approaches
- Certifying or listing trusted age assurance technology providers
- Regularly reviewing methods of age assurance for performance
- Introducing transparency requirements

Keeping with the highest possible levels of data protection & minimisation

The Yoti platform is designed so that all user data is encrypted with a key on the user's phone, and can only be read or shared by the user themselves. No Yoti staff have the ability to decrypt the data, and we cannot mine or sell user data to third parties.

When a user uses a Yoti service to prove their age online, the only information we pass on to the site or vendor is that that person is above or below a given age, e.g. 18 years of age, or that they are of a certain age bracket, such as the 13 to 17 years bracket. We would like to provide some examples of best practices around the storage of data.

Examples of data that Yoti does not store:

- Biometric data
- Text or images of an individual's face or ID document
- Device data, such as an IP address or browser information



Examples of data Yoti stores:

- A shared ID to determine if a user has verified their age before.
- The method of age verification used
- The type of liveness checks performed
- The type of authenticity checks performed
- The time the check took place
- A unique ID that provides an audit of internal decisions performed to produce a token (no images or text).

We are strong believers in data minimisation and user privacy, and that we believe it is best for VSPS providers to be able to access as little information about their users besides what is required as possible.

We believe there is no justification for a site to request to know which city or a month a person was born in, or which citizenship they hold. All they need and should be able to know is that the user is of the age they require their users to be in order to access content or a service.

This is critical for trust, on which the whole AVSMD regime will rely. If users do not feel that they can trust that they remain in control of their own personal information and that it is only used in ways that are consented and secure, they will find ways to go around age assurance measures and trust in the regime and the regulator will suffer.

Enabling choice and inclusion of age assurance approaches

We believe that it is important to offer consumers a range of options to prove age. Our experience shows that consumers prefer to use estimation, or non document based options in over 80% of instances; this may be as in certain use cases an individual does not feel comfortable using a document based approach, because of convenience, lack of access or not owning a document. It is worth recalling that over 1 billion people on the planet¹ do not own or have access to government issued identity documents.

Certifying or listing trusted age assurance technology providers

¹ 'Why ID matters for development', World Bank website, <https://id4d.worldbank.org/guide/why-id-matters-development#:~:text=As%20of%202018%2C%20the%20ID4D,not%20have%20basic%20identity%20documents>



As previously stated, Yoti's age verification technology has been fully or partially audited by a number of trusted, independent third party organisations such as:

- the Information Commissioner's Office (ICO), British Board of Film Classification (BBFC) via NCC Group, the Age Check Certification Scheme (ACCS), Digital Identity Systems Certification (DISC) in the United Kingdom.
- the Association for Voluntary Self-Regulation of Digital Media Service Providers (FSM) and Commission for the Protection of Minors in the Media (KJM) in Germany, listed on the KJM 'Raster'. The KJM Raster provides a clear list of age assurance approaches which it has reviewed and deems suitable for use in the German market. This provides clarity for relying parties, age assurance providers, civil society and consumers alike.

In December 2021, the Information Commissioner's Office published an [Age Appropriate Design Code Audit Report](#)² which rated Yoti as having a 'High' level of age assurance rating. It stated that Yoti offers 'a high level of assurance that processes and procedures are in place, that the organisation is in conformance with the Age Appropriate Design Code and are delivering data protection compliance.'

Following on those two examples, the Media Commission could decide to create a register of trusted age verification technology providers similar to or on the same page as the register of the services that will have been designated under the VSPS regime.

We also believe that a healthy and independent network of trusted third party auditors could be a solution to ensuring that VSPS providers can be assessed to ensure their compliance with transparency and accessibility requirements. This would also allow for the burden of auditing thousands of sites to be shared between the Media Commission and others.

Regularly reviewing methods of age assurance for performance

There could be a requirement for age verification providers to demonstrate their compliance with standards, such as Publicly Available Specification ('PAS') 1296:2018 and the incoming international Institute of Electrical and Electronics Engineers (IEEE) and International Organization for Standardization (ISO) age standards, be audited by Irish National Accreditation Board (INAB) accredited bodies, in order to be listed on a trusted age

² 'Yoti Age Appropriate Design Code Audit Report', Information Commissioner's Office (ICO) website, https://ico.org.uk/media/action-weve-taken/audits-and-advisory-visits/4019830/age-appropriate-design-code-yoti-app-audit-report-executive-summary-v1_0.pdf



assurance providers list by the Media Commission, in the same way that 'Qualified Trust Services' are listed publicly³ or that the German KJM lists⁴ services which it has reviewed.

We would recommend that an independent review should also be commissioned or undertaken by the Media Commission into the robustness of each generic approach. For instance a testing body could review the time, skill and cost that are needed to bypass an VSPS' age check system.

We would also suggest that regulation should require the components of an AI age checking service to be independently reviewed for bias, any AI data sets should be reviewed to ensure that they are created from data sets in accordance with GDPR and not created from scraped images and that age estimation data is instantly deleted.

Providing clear explanations of age assurance methods to users

Yoti has been supportive of the work that has been done by regulators to develop the Age Appropriate Design Code, also known as the Children's Code in the UK, as well as initiatives in California and the European Union. We believe ensuring compliance with the Code is an important step to making terms of service and public policy statements clear and accessible to all members of the public.

We believe VSPS providers should provide users with a clear explanation of the age verification methods they offer to keep users safe. We think the public would benefit from the production of materials using accessible and age appropriate language in which they detail how their age verification measures work.

Assessing attack vectors for each age assurance methodology

One of the elephants in the room has been the ease to circumvent parental consent mechanisms or provide tokenistic weak age gating approaches - such as tick boxes or self assertion of age or reliance on second-hand or historic checks without re-authentication or knowledge based checks which can be shared or traded.

If the spirit of the Age Appropriate Design Code is followed, then both regulators and platforms should be required to consider the best interests of the child, and review which approaches are deemed too weak to offer appropriate safeguards. This means that the

³ UK Trusted List, Information Commissioner's Office website, <https://ico.org.uk/for-organisations/guide-to-eidas/uk-trusted-list/>

⁴ KJM-positively rated age verification systems, KJM website, https://www.kjm-online.de/service/pressemitteilungen/meldung?tx_news_pi1%5Bnews%5D=4890&cHash=e45ae6dfeee26fcd23d10c6994b7a9ef

regulator must restrict access to content, which can impair a child's physical, mental or moral development.

In our opinion, the Media Commission should go further in its guidance and specifically mandate VSPS to only allow age verification through the use of methods and documents which are designated by recognised age standards or in the Good Practice Guide 45 as offering an appropriately risk assessed level of assurance.

The same tools that tech companies are offering to platforms can also be employed by regulators to audit the effectiveness and ease to circumnavigate age gating approaches. Yoti would be delighted to collaborate with the Media Commission on further developing this guidance.

Introducing transparency requirements

In the case of AI approaches, the Media Commission could make a number of additional requirements, such as by requiring age assurance technology providers to:

- Adopt a consistent measurement approach⁵ and transparently show the accuracy of their algorithms across age, skin tone and gender by providing the rates of false positives, false negatives, true positive, standard deviation, the mean absolute error, and positive predictive value.
- Ethically sourced data sets collected in accordance with GDPR.
- Require independent reviews of bias

And in the case of all age assurance approaches, the Commission could also consider making the following requirements:

- Mandate the use of Plain English, so the demographic using the technology can understand the approach used, the terms and the privacy policy, following the United Nations Children's Fund (Unicef) Policy guidance on AI for children.⁶

⁵ Measurement of Age Assurance Technologies, Part 2 – Current and short-term capability of a range of Age Assurance measures, Digital Regulation Cooperation Forum (DRCF) website, https://www.drcf.org.uk/_data/assets/pdf_file/0029/266618/Measurement-of-Age-Assurance-Technologies-Part-2-Analysis.pdf

⁶ Policy guidance on AI for children, United Nations International Children's Emergency Fund (UNICEF) website, <https://www.unicef.org/globalinsight/media/1171/file/UNICEF-Global-Insight-policy-guidance-AI-children-draft-1.0-2020.pdf>



- Ask that age assurance technology be standards-based, such as by asking providers to meet PAS 1296:2018⁷, as the current standard. In future there will be a transition path to the upcoming ISO and IEEE Age Checking Standards.
- Participate in benchmarking, where this service is or becomes available, such as via the National Institute of Standards and Technology (NIST)⁸ which commences a regular programme of global benchmarking of facial age estimation from September 2023.

More information on Yoti's age assurance technologies

Users can perform an age verification using the Yoti reusable digital identity app, which allows individuals to share verified information about themselves on a granular basis. This can also be done via Yoti's 'embedded' services which allow organisations to add a fully integrated identity verification flow into their website or app. It could also be using Yoti's authentication algorithms such as age estimation. These verification options can be integrated as standalone solutions, or via the Yoti age verification portal offering more choice to the end users and configuration options to organisations.

In all verification scenarios, Yoti calculates if the user meets the minimum age requirement to access the website.

If the Yoti reusable digital identity app is used, the user scans a Yoti QR code with the Yoti app to share their age attribute. Yoti then generates a hashed age token, which tells the website that the user is over the required age. The token and Yoti's record of the individual's age, or characteristic as over an age threshold, only last for the browsing session and do not identify the individual personally. Further, no personal information is shared with the adult site beyond the age attribute, making this a private and secure solution. The user's interaction with the website itself remains entirely anonymous.

If Yoti's fully integrated identity verification solution is used, the end user scans or uploads their identity document straight from their web browser or mobile app. An age is computed from the date of birth included in their identity document, and used to establish whether the person is old enough to pass the age verification test.

With Yoti's age estimation solution, users simply look into their phone's camera or their computer's webcam, and Yoti Age Scan will estimate their age. The image is captured and securely transmitted to Yoti's server using 256-bit encryption. Then, Yoti's algorithm gives a result in approximately 1 second. The image is immediately deleted from Yoti's servers and

⁷ PAS 1296, Age Check Certification Scheme (ACCS) website, <https://www.accscheme.com/services/age-assurance/pas-1296>

⁸ Face Analysis Technology Evaluation (FATE) Age Estimation, National Institute of Standards and Technology (NIST) website, https://pages.nist.gov/frvt/html/frvt_age_estimation.html



no record of the user is retained. The only output is an anonymous, hashed age token, used to determine if they are old enough to access the age-restricted content material.

Yoti has also been part of the [EU Consent project](#) devising pan-European interoperable infrastructure for age verification and parental consent. Yoti has a strong ethical focus, having an internal Ethics & Trust Committee and being a certified B Corporation. More on Yoti's approach to privacy, ethical oversight and accuracy can be found in [Yoti's March 2023 Age Estimation White Paper](#). The below is an extract from the White Paper:

How accurate is facial age estimation?

When presented with a clear facial image, our technology compares very favourably with human abilities. Anyone who has used a complicated computer spreadsheet will recognise that in some areas, computers are better than humans at doing some things.

Humans tend to systematically underestimate the ages of older people, and overestimate the age of younger people, and as we ourselves get older, our ability to estimate accurately tends to decrease. When viewing a succession of faces, a person's judgement tends to be influenced by the faces they have just seen - this isn't a problem that affects facial age estimation. These problems clearly have particular implications for provision of age-restricted goods and services, where we need to check whether teenagers are above or below a required legal age.

Currently, the MAE across the entire data set, de-skewed to give equal weighting to male and female subjects for all 65 year olds, is 2.9 years and just 1.4 for 13-17 year olds. Further detail on our algorithm's accuracy, broken down by gender, skin tone and each year of age, is presented in this paper's appendix.

The vast majority of organisations who need to check age need to check whether individuals are over the age of 13, 18 or 21. However, there are additional very important requirements for checking individuals are under an age of interest. We recognise that we still have further to go to reduce bias for older age groups, particularly individuals with skin tone V & VI. However, these older individuals are not materially disadvantaged when the age of interest is for example 18 or 21 and the thresholds are usually 25 or 30 respectively.

About the 'Mean Absolute Error':

Yoti facial age estimation can make both positive and negative errors when estimating age (that is, it can estimate too high, or it can estimate too low). By taking 'absolute' values of each error we mean ignoring whether the error is positive or negative, simply taking the numerical size of the error. We then take the average (or 'arithmetic mean') of all those absolute error values, producing an overall 'MAE'.

The average MAE can be measured as:

- i) the average of each year's MAE - eg. there are 65 year MAEs in the 6-70 age range,
- ii) the average of each age MAE - all age ranges are shown on pages 5-7,
- ii) the average of all the images in the training data (but this data may be skewed towards certain ages with more training data).

There is a clear need for regulators to ensure MAEs are measured independently and consistently to ensure trust in the accuracy of models.

Our various age verification applications have been widely deployed and highly scalable. The age verification industry is now mature and well established, and its contribution to the global digital economy is significant. We provide age assurance technology to major and small content providers alike across the world.

Based on our experience, setting up verification processes on a website or platform can take from as little as 2 hours to one day (approximately 8 hours). Effective automatic verification processes do not require any input from staff. The entire process is quick, streamlined and automated. At scales suggested in points below (millions of verifications), the cost per verification is likely to be closer to a few pennies.

For a one time, account based, age check, Yoti's rate card in a country like Ireland, for the majority of age checks is €0.29. Volume discounts are available. For anonymous, free to view, adult sites, based on repeat daily visits, we charge an annual licence for the combination of offering Yoti facial age estimation, age over response from a government issued document, and a token approach. Yoti digital ID (Yoti app) is offered in conjunction 'free of charge'

To date (August 2023), Yoti operated at large scale and performed over 600 million age checks on behalf of a wide variety of clients located across the world. That is over 450,000



checks per day, or close to the whole population of Cork being age checked twice every day. These included adult content providers, large retailers, gaming sites and platforms.

As stated above, we are a provider of age assurance and age verification tools to a large range of in person and online service and content providers such as video-sharing platform services (VSPS) providers.

Our [March 2023 Yoti Age Estimation White Paper](#), which is available to the public and hosted on our website, provides more information about how technology works. This includes information about our age estimation technology, as well as its accuracy levels, our commitment to the ethical use of our technology, and how we work to train our dataset to make it more inclusive.

We of course would be delighted to provide more information on how our technology works should the Media Commission require it. As part of this consultation response, we would like to make the following recommendations and bring the Commission's attention to the following principles:

As stated previously, we will continue to support efforts by the Media Commission to get the transposition and enforcement of the EU's Audiovisual Media Services Directive (AVSMD) right, and ensure that people are duly protected via age appropriate design. We agree that the Media Commission should focus in particular on age assurance technology and its potential on helping it achieve the aims of the Digital Services Act (DSA) and the Digital Market Act (DMA). We remain fully available to offer any additional supporting evidence, oral or written, that may aid the Media Commission achieve its objectives following the conclusion of this consultation and in the future.

What evidence is there about the effectiveness of age estimation techniques?

As said previously, we welcome and remain supportive of the Republic of Ireland's aim to drive the implementation of robust age assurance to enable age appropriate design and protect minors from harm online. The eyes of the world will be on the Media Commission in terms of its transposition and enforcement of the European Union's Audiovisual Media Services Directive (AVSMD). Indeed, the Media Commission's education and enforcement record will set the tone for platforms and others who will be in scope of the Digital Services Act (DSA) and the Digital Market Act (DMA), and will determine whether these actors will voluntarily move to compliance, or whether the Media Commission will have to use enforcement powers.

Yoti is currently at the forefront of global age assurance policy & standards development across the world, working across continents to shape the future of age assurance.



We are in particular involved in the *Laboratoire pour la protection de l'enfance en ligne*, which is an initiative led by the presidency of the French Republic and the office of the Prime Minister of New Zealand, and which sits within the Christchurch Call framework. This initiative aims to support the sandboxing of age verification technologies, and Yoti is taking part in this sandbox with its partner Meta, and will soon expand this to multiple other global platforms.

As part of this initiative, Yoti provided its age estimation technology to the mobile application Instagram to verify the age of users attempting to change their date of birth from under to over 18. When prompted to choose a verification method between Yoti's age estimation method and an identity document-based verification, **81% of users chose facial age estimation**. This is also the preferred method of age verification by users of the French social media platform Yubo.

We are also involved in an industry-led working group to define a future standard for age verification techniques led by the *Association Française de Normalisation* (AFNOR)⁹, which aims to first apply to French sites, but ambitions to become the basis for a future European standard agreed by all Member States. We are also involved with solution providers to take part in the *Commission nationale de l'informatique et des libertés* (CNIL)'s double anonymity model of age verification trials.

We would like to challenge the following statement: *'More robust age verification involves relying on documents such as a driver's licence or passport'*. For instance, the statement can be qualified by saying that robust age verification with documents requires the following three elements: a document review, face matching and liveness detection.

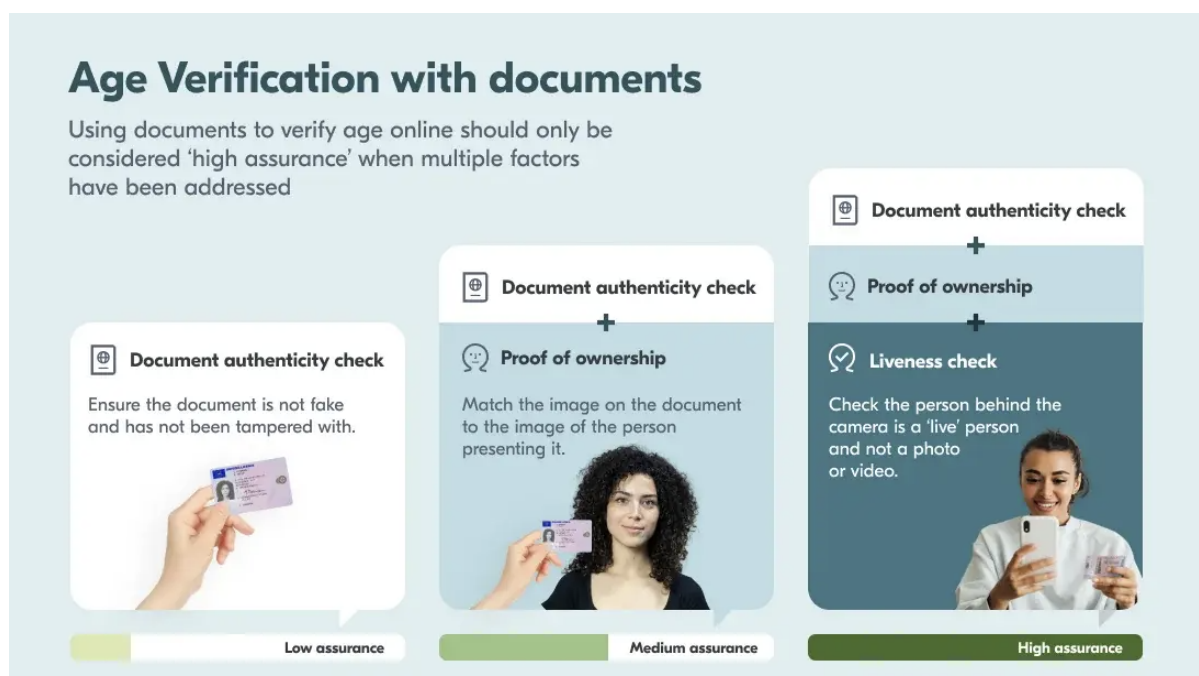
It is commonly known that fake documents are widely used by individuals (often under 18) who wish to access 18 plus goods and services. They are also used by adults who wish to conceal their identity. High numbers of high quality fake identity documents with holograms are available online for just €10 to €60¹⁰ in every Member State and beyond. Hence, it is important to underscore the imperfections of identity document-based age verification methods reliant on identity documents. These approaches fall short of absolute accuracy and are not consistently more reliable than age estimation techniques unless robust authenticity, liveness and face matching is undertaken.

⁹ The member body for France at the International Organisation for Standardisation (ISO), equivalent to Ireland's National Standards Authority of Ireland (NSAI).

¹⁰ [Fakelidentification.co.uk](https://fakeidentification.co.uk/), provided as an example of a site offering fake documents at a low price.

Statistics are rarely cited, but the 2016 European Commission Action plan to strengthen the European response to travel document fraud¹¹ mentions ‘impostor fraud and the fraudulent obtaining of genuine documents increased by 4% and 76% respectively between the first quarter of 2015 and the first quarter of 2016’. There are also 99 million lost and stolen documents recorded by Interpol¹². If the document authenticity check is not complemented by a liveness and a face match check, there is a high risk of impersonation.

It is too simplistic to say that all methods of verification based on identity documents are more ‘robust’ or offer a higher level of assurance, and that all methods of estimation offer a lower level. The level of confidence has to be assessed based on the detailed mechanics of each approach and the ease, or difficulty, for a bad actor to ‘spoof’ the verification and succeed in gaining an incorrect year of age, or incorrectly passing an over or under age threshold.

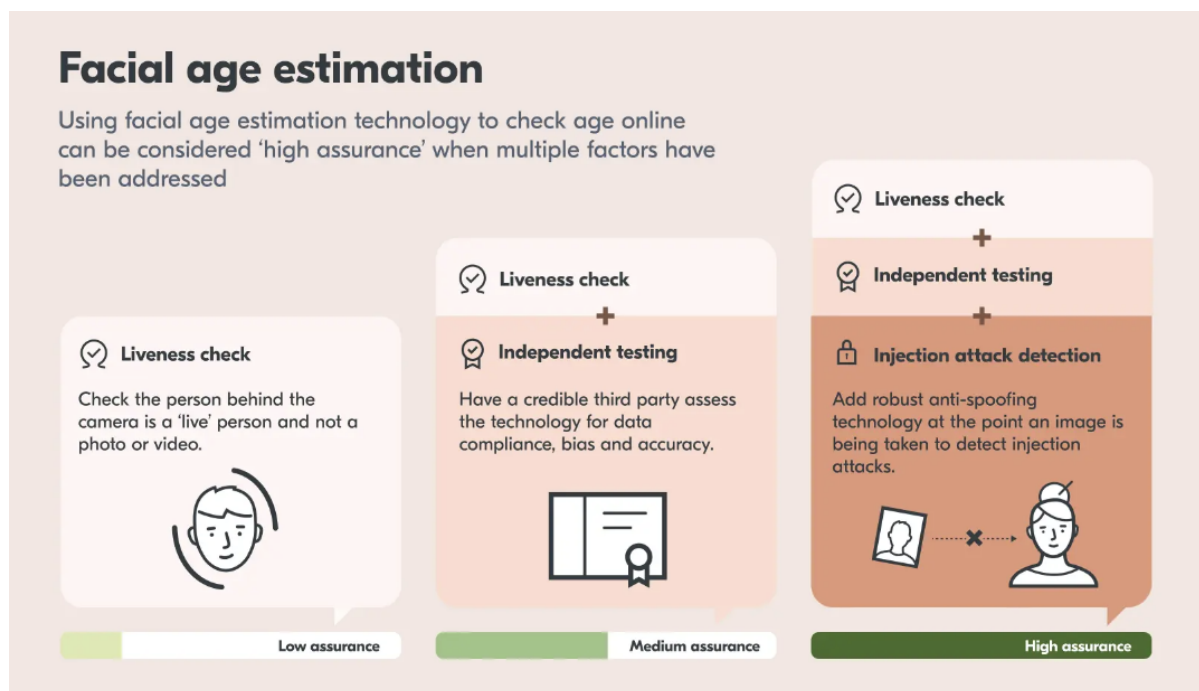


When a liveness test and facial age estimation of adults is undertaken with a threshold of, for instance, 35 years, then there is a 0% chance of estimating a 14-17 old as over 35. Hence it is too simplistic and indeed scientifically naive to claim that a physical ID or database check will always be better than age estimation. The Age Check Certification Scheme (ACCS), an independent testing company, assessed the model and published its certification

¹¹ Communication from the Commission to the European Parliament and the Council, Action plan to strengthen the European response to travel document fraud, EUR-Lex website, <https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52016DC0790>

¹² Interpol stolen and lost travel and identity documents (SLTD) database, Interpol website, <https://www.interpol.int/en/How-we-work/Databases/SLTD-database-travel-and-identity-documents>

report¹³ for Yoti's Age Estimation in November 2020 which concluded that "the system is fit for deployment in a Challenge 25 policy area and is at least 98.89% reliable".



Our True Positive Rate (TPR)¹⁴ for 13-17 year olds being correctly estimated as under 25 is 99.93% and there is no discernible bias across genders or skin tones. The TPRs for females and males 13-17 year olds are 99.90% and 99.94% respectively. The TPRs for skin tones 1, 2 and 3 are 99.93%, 99.89% and 99.92% respectively. This gives regulators globally a very high level of confidence that children will not be able to access adult content.

In addition, our TPR for 6-11 year olds being correctly estimated as under 13 is 98.35%. The TPRs for female and male 6-11 year olds are 98.00% and 98.71% respectively. For this age range, the TPRs for skin tone 1, 2 and 3 are 97.88%, 99.24% and 98.18% respectively so there is no material bias in this age group either.

Once a sensible buffer is established, there is clear evidence that certain types of age estimation, can provide as high or a higher level of confidence in the age of a user than identity documents-based methods of age verification. Hence we feel strongly that the science and evidence needs to be reviewed for each method and that this should be reflected in the future Online Safety Codes.

¹³ Yoti Scope of certification document, Age Check Certifications Scheme (ACCS) website, <https://www.accscheme.com/registry/yoti-ltd>

¹⁴ The True Positive Rate represents the probability that an actual positive will test positive, such as an 18 year old is correctly estimated to be under 25.

The table below shows the false positive rates for a selection of thresholds, for an age of interest of 18. This is an extract from Yoti's March 2023 White Paper¹⁵.

						Average False Positive Rate (weighted equally for each age)
		14	15	16	17	
Test Sample Size		3,167	7,445	10,105	10,214	
Thresholds (years)	20	0.32%	0.67%	1.67%	4.91%	1.89%
	21	0.19%	0.38%	1.02%	2.90%	1.12%
	22	0.16%	0.27%	0.62%	1.57%	0.65%
	23	0.09%	0.13%	0.29%	0.98%	0.37%
	24	0.06%	0.08%	0.25%	0.46%	0.19%
	25	0.00%	0.04%	0.06%	0.27%	0.09%
	26	0.00%	0.00%	0.03%	0.17%	0.05%
	27	0.00%	0.00%	0.01%	0.06%	0.02%
	28	0.00%	0.00%	0.01%	0.06%	0.02%
	29	0.00%	0.00%	0.01%	0.04%	0.01%
	30	0.00%	0.00%	0.00%	0.03%	0.01%

We would also like to offer a clarification in response to the following statement: 'VSPS providers who use this technique have data on the number of accounts that they have detected that belong to minors who have claimed to be adults. But this data is not conclusive in assessing the effectiveness of age-estimation techniques. Data on the proportion of minors that claim to be adults but evade detection would be more useful but is inherently harder to collect.'

Meta have claimed the following in terms of use of Yoti facial age estimation, as cited in the public domain¹⁶, Meta says its age verification tools on Instagram have already had an impact. Since its tests in June, it found that approximately four times as many people were likely to complete its age verification requirements when attempting to edit their date of birth to be over 18. This resulted in "hundreds of thousands" of users being placed in their appropriate age groups. It also claims to have stopped 96% of teens who tried to edit their birthdays to over 18 using these tools. And it said 81% opted for Yoti's video selfie to complete the process.

¹⁵ Yoti March 2023 White Paper, Yoti website:

<https://www.yoti.com/wp-content/uploads/Yoti-Age-Estimation-White-Paper-March-2023.pdf>

¹⁶ 'Facebook to now test age verification tech on Facebook Dating in the US', TechCrunch website, <https://techcrunch.com/2022/12/05/facebook-to-now-test-age-verification-tech-on-facebook-dating-in-the-u-s/>



Superawesome states the following as part of its joint Verifiable Parental Consent (VPC) Children's Online Privacy Protection Rule (COPPA) 2023 application with Yoti and the Entertainment Software Ratings Board (ESRB): *'Facial Age Estimation is effective. On average 35% of users in the EU and U.K. who attempt to prove they are adults are rejected for being under the threshold age. Facial Age Estimation is easier to use – 91%+ of users who start the face scan flow complete it, as compared to 65% for payment cards, or 66% for SSN.'*

Yoti can also produce reports which feature more data than a binary 'fail' or 'pass' result, such as data on the uncertainty value and failure reasons where a user's age could not be provided.

Together with the mean absolute error rate, this type of data and the ability to compare between age assurance methods can enable sites to assess the effectiveness of age estimation techniques, as well as completing their own underage access risk assessments, which form the backbone of several online safety regimes across the world.

Some examples of the data that can be analysed to assess the performance of age estimation techniques include:

- The number of checks completed
- The number of checks that were completed and which returned an age result, as well as the detail of why checks that did not return an age failed, for instance if the age check failed because the liveness test was not passed, or because of poor quality submissions.
- The number of checks attempted, which is the number of sessions where the user has attempted at least one age estimation scan
- The age distribution for completed checks, which is the number of checks performed for each estimated year of age.

Finally, age estimation is also a means to combat social exclusion for the significant numbers of individuals around the world who do not possess a state-issued photo ID document. These documents (such as passports and driving licences) can be expensive to apply for and obtain, and a significant proportion of young people do not possess them. Large numbers of physical ID documents are also lost every year¹⁷, increasing the risk of identity fraud as well as incurring a replacement cost. Facial age estimation is designed with user privacy and data minimisation in mind. It does not require users to register with a

¹⁷ 'Drivers lose almost a million licences in the last year', UK Government website, <https://www.gov.uk/government/news/drivers-lose-almost-a-million-licences-in-the-last-year>



provider such as Yoti, or provide any documents to prove their identity. It is unable to personally identify an individual, it simply estimates a person's age from analysing their face.

MODEL CODE

A reference model for regulatory or self regulatory approaches to harm reduction on social media

January 2023

Introduction

Carnegie UK's [work on social media regulation](#) has underpinned policy formation in the UK. In 2018 we described a new way of reducing harm arising from social media that recognised international human rights. We proposed regulation that focussed not on individual pieces of content but on the systems and processes that companies use to run the business, and to incentivise and distribute content. In the UK that took the form of a statutory duty of care with an independent regulator; in other regimes, such as the DSA, a due diligence process. Since 2018 we have developed our model with victim groups, regulators, governments, parliamentarians, social media companies, civil rights organisations and law enforcement. Our approach has been discussed in the UK, the EU, the G7, the United Nations and in places as far apart as Canada, Germany, Mongolia, Slovakia, Australia and New Zealand.

Carnegie UK has argued that:

- the scale, speed and variety of content on internet platforms make it difficult to regulate content directly;
- that design features, business model and user tools have an indirect impact on content and that platforms have some responsibility for those choices;
- these features can have an affect at each of a number of stages in the content distribution chain (seen as a four-stage model, below);
- a regulatory approach focussed on risk assessment of these features may improve the content environment without excessive reliance on content-removal, and balance the potentially conflicting rights of users through proportionate risk mitigation strategies;
- while individual content domains may be differently affected by particular design features, and different user tools may be helpful depending on context, all content flows through the same distribution chain in a service (so the same questions arise about risk of features, even if the answer might differ across content domains).

On this basis, it seems that a common framework could be developed by reference to the four stage information flow model. The framework would form the basis for a company approach to risk assessment and mitigation. This framework could be deployed across multiple content domains and jurisdictions. In adopting this cross-cutting approach, design-based risk mitigation measures can be seen to have cross-domain – and cross harm - effects. The approach may therefore be more efficient for service providers in tackling specific harms across a range of content domains and -potentially – across jurisdictions.

This approach, which sets down principles rather than detailed rules, is flexible:

- defining a skeleton approach allows a company to develop and apply the framework within its own context (rather than imposing specific technical answers across the

sector) – this may allow for a providers to compete on the basis of their approach to safety, allowing greater choice to users;

- it is future-proofed for similar reasons;
- it allows modular development – so that content domains may be incorporated or not, depending on service provider and requirements of local jurisdiction;
- it can sit as part of formal, legal regulation, be part of self-regulatory initiatives, or sit against international framing and provide a common thread amongst these multiple legal structures (though these different legal structures may be more or less efficient in terms of impact and enforceability).

The model code has many similarities with the UN Guiding Principles on Business and Human Rights (Ruggie) and the OECD Guidance for Multinational Enterprises and would be consistent with those approaches.

The four stage information flow model, which reflects the role of the platforms in creating and influencing the flow of content from their users, comprises the following:

- access to the service and content creation;
- discovery and navigation;
- user response tools; and
- platform response.

Access to the service and content creation includes tools available to users to create content (e.g. filters, nudification apps and mechanisms for labelling content), as well as restrictions (eg limits on frequency of posting) but also includes the user sign-up process and the terms of service for use of the platform. So questions around anonymity, multiple accounts, the acceptability of bot accounts and disposable accounts could all be considered here as well as the adequacy of the terms of service (assessed either against national law or international law standards, as appropriate). The main focus in community standards or terms of service tends to lie on user-facing provisions; advertising content policies should not, however, be forgotten; nor the impact of advertising revenue sharing business models on user content creation.

Discovery and navigation covers all sorts of recommendation tools, and features for organising content such as hashtags and feeds highlighting trending issues, as well as search functions/autocomplete. Advertising delivery systems also fit here, including advertiser sign-up processes (KYC), ad content policy and audience segmentation tools.

User response tools allow the user to curate and adapt the online environment, but this category also includes tools for engaging with content (like buttons for example, or features to facilitate reposting and sharing) as well as the ease of making complaints.

Platform response includes moderation and complaints processes, including any user rights of appeal, crisis protocols and transparency reporting.

At each of the four stages, an intervention could be any one of: an ex ante design choice; the provision of tools or other mechanisms; or content specific responses. For example in terms of discovery a service could choose to optimise for authoritative sources; allow users more control to curate their own feed; or introduce suppression measures related to particular content or speaker.

The reference model

This model code is drafted to sit along international human rights standards as a self regulatory tool. As such it is not phrased, in the main, in mandatory terms. A variant of this code, sitting within national law could be more prescriptive, specifying clearly mandatory requirements. Nonetheless, within this model code there are some principles that are phrased in mandatory language. The approach is based on risk assessment and mitigation so a risk assessment approach itself is mandatory. Further principles identify issues which in our opinion should be

considered in a risk assessment. We also give examples of specific features which are often considered risky.

The nature of this high level, principles based approach means that mitigations cannot be identified here – they will be context specific. An exception to this are steps central to the protection of user rights and which seem relevant no matter the content domain and jurisdiction. Such universal mitigating steps we have expressed to be a mandatory requirement, though the details of implementation may be informed by the risk assessment.

There is not a perfect answer to social media regulation that will fit all countries. A race for a commonality can quickly plummet to only the lowest common denominator which exposes the international disparity in application of the ICCPR leaving many people vulnerable. A modular approach (after Watkins, Ness) using a reference model as set out here could be a basis for international agreement based on common principles. This can provide both help for countries with few resources and the beginnings of a detailed framework for those equipped to regulate.

Professor Lorna Woods OBE, University of Essex
William Perrin OBE, Trustee Carnegie UK
Maeve Walsh, Associate Carnegie UK
January 2023

Contact: 

Company orientation towards reduction of harm

Principle 1: Responsibility, Risk Assessment, Mitigation and Remediation

1. The service provider must have a policy commitment to seek to reduce harm, whether in general or in relation to a set of harms, arising from the operation of their service endorsed by the board and significant subcontractors.
2. The governing board of the service provider should apply the United Nations Guiding Principles on Business and Human Rights, as should significant subcontractors in the supply chain. Large multinational service providers with complex supply chains should comply with the OECD Guidelines for Multinational Enterprises. Particular regard should be paid to risks of harm to media and democratic freedoms.
3. As a foundation for harm reduction activity, service providers must carry out a suitable and sufficient risk assessment of their entire service, including risks arising from the practice of outsourcing responsibilities. In doing so, service providers should engage with relevant experts and organisations representing groups adversely affected by operation of the service. A suitable risk assessment will follow international standards if available or best practice. It shall cover all territories where the service has a non-trivial user base, reflecting their local circumstances accordingly.
4. Where harm reduction is focussed on one or a few specific content domains (eg violence against women and girls), service providers should include a survey of the extent to which the relevant content domain arises and results in harm on its service.
5. Risk assessment must also be carried out in relation to the launch of any new service or new feature. Providers of high harm or very large services should operate a precautionary principle in introducing new features, only gradually increasing their availability while monitoring for harm in dialogue with representatives of victims.
6. The service provider must produce a risk mitigation plan addressing the issues raised in the risk assessment and at least covering the issues covered later in this code (including product testing).
7. The service provider must identify appropriate metrics to assess the appropriateness and success of the mitigation plan (or any part thereof) and use them to assess the effectiveness of the mitigation plan regularly (at least annually) and to update it as appropriate. Metrics should be designed so as to allow comparability across assessment periods.
8. The service provider must remain vigilant at all times to reasonably foreseeable events that could give rise to significant harm such as elections, festivals, sports matches etc or observable yet unforeseen events such as civil unrest, war or severe ethnic tensions and mitigate the harm arising from their services in these contexts. Advances in technology leading to or exacerbating harm will occur and should be mitigated as part of ongoing vigilance. In general the service provider should review the success of the mitigation plan at least annually and revise the plan as appropriate.
9. In reviewing progress, service providers must engage with relevant experts and organisations representing groups adversely affected by the relevant content.
10. Risk assessments and mitigation plans should be recorded, retained for not less than three years and published on the service provider's website in an accessible manner in languages commonly used on the service. The service provider should consider instructing third party audits from independent appropriately qualified auditors.

11. Risks assessments should not assume that users and the way they respond to the service and the content on the service are homogenous. Risk assessments must take into account the characteristics of different groups and the differential impact of the features on them as well as the specific risk of harm to which they are exposed. Specifically, harms arising for children should have a separate risk assessment and mitigation process, informed by General Comment No. 25 of the UN Committee on the Rights of the Child in relation to the digital environment.
12. This first principle is a foundation for principles 2-12. The following principles are applied with reference to assessing and mitigating risk arising from the operation of the services, in all non-trivial geographic markets and including the actions of significant sub-contractors.

Principle 2: Safety by Design

1. Bearing in mind the outcomes of the risk assessment, the service provider must implement appropriate technical and organisational measures to embed safety by design in the development and the running of the service and its features. Safety by design does not mean the elimination of all risks but rather the inculcation of an approach where appropriate choices about understanding, minimising or allocating risk can be made.
2. The service provider must take steps to ensure that the design process takes into account the different characteristics of users, aiming to design inclusively.
3. As part of its risk assessment and mitigation processes, the service provider should carry out or arrange for the carrying out of such testing and examination of its service and business systems (including any advertising or paid content systems) to assess the safety of the service by reference to the harms caused in each relevant content domain by the operation of the service. Testing should include "abusability testing" as well as identifying whether features and tools scale well from a viewpoint of user safety.
4. Testing should include systems and tools for recommendation, content curation and moderation, especially automated tools, but also including user empowerment tools. The service provider should test tools provided by third parties, or ensure that those tools have been adequately tested for safety in the service provider's particular context.
5. At least annually, but informed by the risk assessment process in Principle One and testing, the service provider must review and, where indicated by the review, revise those technical and organisational safety by design measures and/or tools. In reviewing design features, the service provider shall consult with relevantly qualified external experts where appropriate.

Principle 3: Education and Training

1. The service provider must put in place appropriate, updated education and training on harm reduction for all staff and subcontractors involved in the content production and distribution chain. This includes senior executives, designers, developers, engineers, customer support and moderators.
2. Where possible the training should be designed in consultation with independent trusted flaggers and/or representatives of survivors of online harms so as to ensure diversity and inclusion.

3. The service provider should provide staff in section 3.1 above with relevant information, training and support on human rights including the importance of an independent media and democratic voice.

Principle 4: Supply Chain Issues

1. The service provider which outsources any part of its business, including moderation of content, applications, GIFs, images or any other content or tools, including 'safety tech', should ensure that each vendor adheres to safety principles and processes in order to deliver the service provider's Terms of Service or Community Standards.
2. Reliance on outsourced content, features or functions must be included as a fact in the risk assessment and mitigation strategy.
3. People doing outsourced work (such as content moderators) should be protected from reasonably foreseeable harm arising from their task through amongst other things a human rights due diligence process such as that described in the OECD Guidance for Multinational Enterprises.

Access to platform and creation of content

Principle 5: Access to the Service

1. The service provider should ensure, and be able to demonstrate, that its sign-up processes take an appropriate, proportionate approach to the principle of "know your client" (KYC), both in relation to users and advertisers. In particular, insofar as the service provider allows anonymity or pseudonymity, these should be included in the risk assessment (taking into account e.g. the user base, focus of the service) and appropriate mitigating steps for any risks identified implemented. Other aspects of that should be taken into account in the risk assessment include:
 1. the extent to which multiple accounts are permitted
 2. bot accounts
 3. extent to which lack of friction in account creation and availability of multiple accounts allow for 'disposable accounts'.
2. The service provider must make its terms of service (including any privacy policy) and/or community standards visible to would-be users and advertisers before they sign up to the service. The terms of service and/or community standards must be expressed in clear and easy to understand language bearing in mind the comprehension capabilities of groups likely to use the service. This includes providing different language versions of the terms of service and/or community standards appropriate to the territories in which the service is made available. The service provider should have in place expanded guidance explaining their terms of service/privacy policies/community standards (and how these are developed, enforced and reviewed, plus the role of relevant survivors' groups and civil society in developing them). It should ensure that training and awareness tools are readily available to users on the Terms of Service and Community Guidelines to ensure users are aware of permitted content and behaviours on the platforms.
3. A service provider must have terms of service and/or community standards in respect of its users that are fit for purpose taken against its values, local laws and international human rights.
4. A service provider must undertake regular systemic reviews of its terms of service and/or community standards to ensure that they remain up-to-date, effective and proportionate, and amend them when appropriate (taking into account the findings under Principle 1)

5. To allow a user to make an informed choice when deciding whether to use a service, the Terms of Service should clearly state the risks of harm identified in the risk assessments and the steps taken to mitigate them, including if no steps are taken.
6. A service provider should prompt its users to consider their safety and privacy settings; these features are to be designed appropriately in the light of the risks present on the service. The system must default to the most secure settings.

Principle 6: Creation of Content

1. A service provider should consider the appropriate levels of friction in the content-posting process in the light of its risk assessment – for example prompts about harmful language used in a post; number of posts permitted over a given time period; provision of content wrapper features; more than one click required to repost content.
2. A service provider should consider whether any monetisation or revenue-sharing arrangements with content providers provide incentives for or provide financial support to harmful content, and take appropriate steps to mitigate any such risk.
3. A service provider should include any tools it provides for the creation of content in its risk assessment – this includes but is not limited to bots (including chatbots), bot networks, deepfake or audiovisual manipulation capability, the ability to embed content from other platforms and synthetic features such as GIFs, emojis and hashtags. It should consider implementing oversight on third party tools that it allows to interact with its service.

Discovery and navigation

Principle 7: Discovery

1. The service provider should review their recommender systems, whether in relation to content or to other users to follow, especially their automated systems, so they do not promote harmful content in general or that related to a specific content domain identified as problematic. The service provider should check automated systems for bias (e.g. arising from training data). The service provider should consider the risks of tools/features used for organising content (eg hashtags) and what safeguards should surround their use, for example to prevent terms inciting violence against minoritised groups being used.
2. The service provider should consider the impact of autoplay functions, especially in the context of content curated or recommended by the provider. When a service provider seeks to take control of content input away from the person in this way the provider should consider how this feature might affect a person's right to receive or impart ideas.
3. The service provider should consider whether to provide appropriate information to its users about the accuracy (or otherwise) of information (eg flagging content that has been fact-checked) and should make its policies in this regard available.
4. The service provider must consider how its advertising delivery systems affect content seen by users. In particular, it must consider the circumstances in which targeted advertising may be used and managerial oversight over the characteristics by which audiences are segmented where those segments might be computer or user - generated.

5. The service provider must have terms of service and/or community standards in respect of its advertisers that are fit for purpose taken against its values, local laws and international human rights and should have processes in place to enforce that policy consistently.
6. The service provider must consider the need for explainability or interpretability, accountability and auditability in designing AI/ML systems.
7. For users who are children, the service provider should ensure that Principle 7 is applied to reflect their particular characteristics and vulnerabilities, including their right not to see some information.

Principle 8: Navigation

1. Interface design must adopt a user-centred approach, which takes into account an appropriate (bearing in mind the user base of the service) level of safety; interface design must not manipulate users (no dark patterns).
2. The service provider should consider the impact of autocomplete functions and have systems in place to oversee the process of suggesting auto-completes.
3. If the service relies on personalisation, it should consider how to institute oversight over the segments used for personalisation and have policies in place to identify unacceptable or unethical labels, such as might emerge through automation.
4. The service provider should consider the risks around embedded content from other services and click through to external sites, especially in relation to advertising content.

User response, user tools

Principle 9: User-empowerment Tools

1. The service provider must consider what tools, in addition to content and behaviour reporting tools, are necessary to allow users to improve their control of their online interactions and to improve their safety. These could include:
 - a) controls over recommendation tools, so a user could choose for example to reject personalisation;
 - b) user-set filters (over words, images, sound, videos or topics);
 - c) tools to limit who can contact/follow a user, or to see a user's posts;
 - d) tools to allow users to block or mute users, or categories of user (eg blocking anonymous and/or unverified accounts);
 - e) Controls for the user over who can and cannot redistribute their content or user name/identity in real time.
2. The service provider should ensure that tools provided following Principle 9(1) are easy to use by all groups of users likely to access the service and take reasonable steps to ensure their prominence such that users are aware they exist.
3. For users who are children, the service provider should ensure that Principle 9 is applied to reflect their particular characteristics and vulnerabilities – in particular 9(1)(c).

Principle 10: Virality

1. The service provider should consider the speed and ease of content transmission. This could include, for example, methods to reduce the velocity of forwarding and therefore the occurrence of harm cross-platform.

2. The service provider should assess the risks posed by any features/tools (eg upvote/down vote; like buttons) provided that encourage users to respond and/or to engage with other user's content.

Principle 11: Reporting and Complaints

1. The service provider must have reporting processes that are fit for purpose, that are clear, visible and easy to use and age-appropriate in design and cover all content and behaviour (whether user-generated, service generated (eg autocompletes) or advertising-based). A service provider should consider whether some forms of complaint (eg harassment; image-based sexual abuse) need specially designed reporting processes.
2. The service provider should allow users and others to complain about unsafe design features that are not 'content'.
3. The service provider must provide the opportunity for non-users who are affected by content or behaviour on the service to report that content and/or behaviour
4. A service provider should record complaints in a sufficiently granular manner to feed into risk assessment review processes. The typology should be developed with survivor representatives.

Platform response

Principle 12: Moderation

1. The service provider's policies must be effectively and consistently enforced in accordance with its detailed policies and further guidance. Such further guidance must be in accordance with national law and international human rights.
2. The service provider must have in place sufficient numbers of moderators, proportionate to the service provider size and growth and to the risk of harm, who are appropriately trained to review harmful and illegal content and who are themselves appropriately supported and safeguarded.
3. Where automated tools are used, the service provider must put in place processes to ensure those tools operate in a non-discriminatory manner and that they are designed in such a way that their decisions are explainable and auditable. Users should be informed of the use of such tools. Machine learning and artificial intelligence tools cannot wholly replace human review and oversight.
4. The service provider must establish clear timeframes or other benchmarks for action against non-compliant content.
5. Action in relation to a complaint must be proportionate to the severity of the harm likely to be caused; content contrary to the criminal law is to be dealt with swiftly. The terms of service should make clearly the nature of any such action and the circumstances in which it would arise, as well as details of any appeals process. Action could include:
 - a) Label content as inaccurate/misleading;
 - b) Demonetise content;
 - c) Suppress content in recommender tools and/or search engines;
 - d) Geo-blocking of content;
 - e) Suspension of content;
 - f) Removal of content;
 - g) Non-recommendation of user and/or group as person to follow;
 - h) The existence of a strike system;
 - i) Geo-blocking of account;

- j) Suspension of account;
 - k) Termination of account.
6. The service provider should have systems of assessment and feedback to the initial reporter and the owner of content that has been flagged and actioned to ensure transparency of decision making. Users should be kept up to date with the progress of their reports and receive clear explanations of decisions taken.
 7. The service provider should consider the risk of abuse of complaints processes and put in place appropriate safeguards. It should put in place a right of appeal on all decisions made concerning illegal or harmful content, or content that has been flagged as illegal or harmful content. This system cannot displace user rights to take action before the courts. All users must be given a right to appeal any measures taken against them, whether in full or in part. Users must be able to present information to advocate their position.
 8. The service provider should have appeals systems which must take no longer than seven days to assess appeals, except in exceptional circumstances which are unforeseeable and beyond the provider's control (see Principle One for discussion of foreseeable events – such as elections – and unforeseeable ones – such as a war).
 9. The social media provider must consider putting in place an appropriate trusted flagger programme that maintains its independence from the service provider and from governments. The service provider should:
 - a) ensure trusted flaggers are not used as a sole provider of flagging content;
 - b) ensure trusted flaggers are appropriately compensated, while not compromising their independence;
 - c) hold regular meetings with members of the trusted flagger programmes to review content decisions and discuss any concerns;
 - d) provide support to trusted flaggers who are exposed to harmful content in line with the service provider's support to its own moderation teams.
 - e) The service provider should have a crisis response protocol that plans for crises of different types in general and, for foreseeable crisis-types, has methodologies to enable the continued delivery of the service without causing harm in accordance with international best practice. This should include occasions when a government seeks to exert undue influence. All protocols should be drafted in clear and precise language. These protocols should include conflict affected and high risk areas, and processes for identifying and monitoring such areas based on existing classifications (eg OECD States of Fragility) as well as monitoring statements from bodies such as the UN or the International Red Cross. Protocols should be tested before deployment and regularly audited in operation.

Principle 13: Survivor Support and Remediation

1. The service provider must take steps to ensure that users who have been exposed to harmful material are directed to, and are able to access, adequate support in the language victims might use. Support can include –
 - a) Signposting and access to websites or helplines dealing with the type of harmful content viewed by the user or witnessed by others who may be affected by the content, even if not the designated target;
 - b) Information from, and contact details for, services providing victim support or mental health support after being exposed to harmful materials;
 - c) Strategies to deal with being exposed to harmful material.

Principle 14: National Law

1. The service provider must have in place a point of contact for law enforcement authorities for each nation in which the service operates. The contact is responsible for

giving information about criminal content to law enforcement authorities in accordance with national law and international human rights (including but not limited to privacy).

2. This information includes –
 - a) Information about the content;
 - b) The details of the user, including location;
 - c) Details of enforcement action on the content undertaken by the provider; and
 - d) Other materials relevant to criminal investigations.
3. Information requested by law enforcement authorities in accordance with the law should be delivered with the time frame specified in relevant law, or, where national law is silent and the time frame is reasonably practicable, in the request.
4. Effective protections should be put in place by the service provider to ensure flagging and court orders are not used for malign purposes to remove content deemed objectionable, by government agencies or law enforcement of any kind, nor powerful individuals which is neither contrary to the law nor to the Terms of Service.

Carnegie UK submission to Coimisiún na Meán call for inputs on “Developing Ireland’s First Binding Online Safety Code for Video-Sharing Platform Services”

1. We welcome the opportunity to respond to the Call for Inputs on the proposals for an online safety code for VSPs. This brief submission covers a copy of our Model Code of Practice which we developed to provide a reference model for regulatory or self-regulatory approaches to harm online. Professor Lorna Woods discussed this model briefly with Niamh Hodnett, after they both took part in a panel on online safety regulation at the FOSI conference in Dublin in June and she suggested that we submit it to the call for input. We would be very happy to talk further to officials at Coimisiún na Meán about it, whether in terms of the general approach or how specific issues are addressed, and will also look out for the consultation on the draft Code later in the year.

About Carnegie UK

2. Carnegie UK’s objective is better wellbeing for people in the UK and Ireland. Over the past five years, we have shaped the debate in the UK on the reduction of online harm through the development of, and advocacy for, a proposal to introduce a statutory duty of care to reduce Online Harms. Our proposal for a risk-management regime within which social media companies design and run safer systems - not for government to regulate individual pieces of content – still underpins the foundations of the Online Safety Bill, which will pass into law in the UK this autumn. The approach is also reflected in the EU’s Digital Services Act. We had a number of meetings with Irish government officials during their policy development work ahead of Ireland’s Online Safety and Media Act. All our work can be found [here](#).
3. Over the past 18 months, we have carried out work – led by Professor Lorna Woods (University of Essex), in conjunction with a number of civil society organisations, academics and other expert groups – to develop principle-based model codes of practice that act at a systemic level to help tech companies assess and reduce the prevalence of online harm on their services, reflecting guidance at international level on corporate social responsibility – notably the [UN Guiding Principles](#) and the work of the OECD. The work started with a [code of practice on hate speech](#), which then informed [ad hoc advice for the UN Special Rapporteur on Minority Issues](#). We then adapted this approach to produce – through the same collaborative process with a coalition of campaigners and academics – a [code of practice on online violence against women and girls](#). This code – and the coalition’s advocacy and campaigning around it – was instrumental in persuading the UK Government to address the lack of protections for women and girls in the Online Safety Bill, bringing in an amendment at recent Lords Report stage to require Ofcom to produce guidance on this issue. We have recently submitted evidence drawing from this work to

the [consultation on the UN Global Digital Compact](#).

4. The codes demonstrate and test the application of underpinning principles. Prof Woods has built on this work, drawing out the underlying risk management principles used in these codes, to develop a universal [Model Code](#) (and attached as an annex) that could inform how a company could approach risk assessment and mitigation and be applied across multiple content domains and jurisdictions. Despite the importance of design to reducing risk of harm, content moderation and response to user complaints remain important. A service provider should be aware of the extent to which it creates risk or is open to abuse, as well as the need to respond to problems arising. We set out some of the key principles behind this below which might aid Coimisiún na Meán officials considering how to approach the Irish code of practice work, with further detail available in the narrative sections within the Model Code itself.

Code of Practice principles

5. Professor Woods' work sets out how a common framework could be developed by reference to a four-stage information flow model (see para 7), forming the basis for a company approach to risk assessment and harm mitigation. This allows interventions to be made at numerous points in the communication distribution chain. It recognises that the causal relationships are complex and that an approach that focusses on one technological fix, or analyses features and functionalities out of context, are unlikely to be helpful. This framework could be deployed across various types of service, multiple content domains and in different jurisdictions. In adopting this cross-cutting approach, design-based risk mitigation measures can be seen to have cross-domain – and cross-harm – effects. The approach may therefore be more efficient for service providers in tackling specific harms across a range of content domains and -potentially – across jurisdictions; in a sense it may allow for interoperability between different jurisdictions. We would suggest that it would be an adaptable starting point for considering the questions set out in 4.1 and 4.2 regarding how flexible an Online Safety Code should be, and how it should be structured.
6. This approach, which sets down principles rather than detailed rules, is flexible:
 - defining a skeleton approach allows a company to develop and apply the framework within its own context (rather than imposing specific technical answers across the sector);
 - future-proofed;
 - allows modular development – so that content domains may be incorporated or not, depending on service provider and requirements of local jurisdiction
 - it can sit as part of formal, legal regulation, be part of self-regulatory initiatives, or sit against international framing and provide a common thread amongst these multiple legal structures (though these different legal structures may be more or less efficient in terms of impact and enforceability). The model code has many similarities with the UN Guiding Principles on Business and Human Rights (Ruggie) and the OECD Guidance for Multinational Enterprises and would be consistent with those approaches.

7. The four-stage information flow model, which reflects the role of the platforms in creating and influencing the flow of content from their users, comprises the following:
 - access to the service and content creation;
 - discovery and navigation;
 - user response tools; and
 - platform response.
8. **Access to the service and content creation** includes tools available to users to create content (e.g. filters (concerns around body dysmorphia for example), emojis (racist abuse) nudification apps – the most usual use of which is to demean women - and mechanisms for labelling content – which can support user empowerment as well as moderation), as well as restrictions (eg limits on frequency of posting) but also includes the user sign-up process and the terms of service for use of the platform. So questions around anonymity, multiple accounts, the acceptability of bot accounts and disposable accounts (which have been raised as concerns in relation to a number of issues including online abuse as well as disinformation) could all be considered here, as well as the adequacy of the terms of service (assessed either against national law or international law standards, as appropriate). It is not sufficient that terms of service mention a topic; the terms of service should be sufficiently granular to address different issues appropriately but also to be clear to users. The main focus in community standards or terms of service tends to lie on user-facing provisions; advertising content policies should not, however, be forgotten; nor the impact of advertising revenue sharing business models on user content creation (clickbait headlines; ‘outrage farming’).
9. **Discovery and navigation** covers all sorts of recommendation tools, and features for organising content such as hashtags and feeds highlighting trending issues (both of which are useful but may support pile ons and harassment), as well as search functions/autocomplete. Anti-Semitic or anti-Muslim autocomplete, for example, have been noted on some of the major search services. The significance of some of these tools has been recognised in the policy process – for example in the DSA’s requirement for recommender system transparency (Art 27 DSA) as well as the limitations on profiling for very large services (Art 38). Advertising delivery systems also fit here, including advertiser sign-up processes (KYC), ad content policy and audience segmentation tools. The risks emerging from LLM-based search functions are also emerging, for example in research by CCDH on [eating disorder content](#) and [harmful misinformation](#).
10. **User response** tools allow the user to curate and adapt the online environment, but this category also includes tools for engaging with content (like buttons for example, or features to facilitate reposting and sharing) as well as the ease of making complaints. The like button and similar features may also provide feedback mechanisms reinforcing behaviours; ease of forwarding can contribute to virality (eg in relation to disinformation). User labelling of posts should also be considered here.

11. **Platform response** includes moderation and complaints processes, including any user rights of appeal, crisis protocols and transparency reporting. The role of trusted flaggers fits here.
12. At each of the four stages, an intervention could be any one of: an *ex ante* design choice; the provision of tools or other mechanisms; or content specific responses. For example, in terms of discovery a service could choose to optimise for authoritative sources; allow users more control to curate their own feed; or introduce suppression measures related to particular content or speaker.

The reference model: applying it to the Irish context

13. This model code is drafted as a self-regulatory tool, but bearing in mind international human rights standards, especially as applied to businesses. As such it is not phrased, in the main, in mandatory terms; it identifies issues for consideration. A variant of this code, sitting within national law could be more prescriptive, specifying clearly mandatory requirements where appropriate. Nonetheless, even within this model code there are some principles that are phrased in mandatory language. The approach is based on risk assessment and mitigation so a risk assessment approach itself is mandatory. Further principles identify issues which in our opinion should be considered in a risk assessment. We also give examples of specific features which are often considered risky. The nature of this high level, principles-based approach means that mitigations cannot be identified here – they will be jurisdiction, platform and problem domain specific. As an exception to the non-mandatory approach steps central to the protection of user rights are relevant no matter the platform, content domain and jurisdiction. Such universal mitigating steps, aimed at protecting users’ rights – including their fundamental rights - are therefore expressed to be mandatory requirements. While the details of implementation may be informed by the risk assessment, there should be a base level standard set of effectiveness, fairness and appropriateness for these processes.
14. This submission does not consider how the specifics set out in the call for input could slot into this model code structure – our capacity, given the current stage of the Online Safety Bill, is limited in that regard and our primary aim here was to ensure that this work was drawn to the attention of the Commission. We would, however, be happy in the early autumn to help officials work through the applicability of this approach to their needs and, vice versa, would welcome the opportunity to test whether the hypotheses above, re the flexibility and adaptability of this model code approach for different jurisdictions, are valid.

August 2023

Carnegie UK

Contact: [REDACTED]



**Ionad
Frithbhulaíochta
Anti-Bullying
Centre**

Coimisiún na Meán, Call For Inputs: Online Safety

Developing Ireland's First Binding Online Safety Code for Video-Sharing Platform Services

Authors: Dr. Sandra Sanmartín Feijóo, Prof. James O'Higgins Norman, Dr. Tijana Milosevic, Dr. Megan Reynolds, Mr. Kanishk Verma, Mr. Derek Laffan, Dr. Darragh McCashin

Thank you for providing us with an opportunity to contribute to this consultation. As advised in the guidelines, we have addressed those questions that we think we can reasonably provide answers to based on our research and scientific evidence that we are aware of, while referring to the relevant literature and publications.

Question 1: What do you think our main priorities and objectives should be in the first binding Online Safety Code for VSPS? What are the main online harms you would like to see it address and why?

It is very difficult for us to answer this question definitively. As a global centre conducting research into prevalence and impact of a wide range of online risks and harms outlined in the Bill, we suggest that the code should develop a set of provisions that would require companies to effectively address all types of risks and harms¹ as required to transpose the AVMSD; and as outlined in the online safety portion of the Online Safety and Media Regulation Bill. This is especially the case if the Commission is aiming to have only one Online Safety Code, rather

¹ For a distinction between risk and harm in the context of children and online safety, please see the following resource from the EU Kids Online Research network: <https://blogs.lse.ac.uk/medialse/2020/02/11/more-online-risks-to-children-but-not-necessarily-more-harm-eu-kids-online-2020-survey/#:~:text=Risks%20and%20harms&text=However%2C%20the%20EU%20Kids%20Online,consequences%20of%20exposure%20to%20risk>

For effective implementation of online safety codes, it would be important to understand that an online risk (e.g., sexting or exposure to pornography) does not always lead to unequivocal harm. This distinction between risk and harm is important when designing provisions that are to balance children's rights to protection on the one hand and participation on VSPS and child privacy, on the other. I.e., over-protection can hamper participation rights, and a certain amount of exposure to risk is necessary for resilience building.

than to develop more codes (e.g., each code would cover one or a set of specific online risks and harms).

Looking at some other examples internationally, The Australian Online Safety Commissioner first requested that the relevant industry association develop a code for class 1A and 1B material, which covers child sexual abuse material and imagery, terrorist content, extreme crime and violence.² Subsequent codes would cover other types of harmful online content. It might be advisable to consider developing separate online safety codes for different types of online harms.

It is our understanding that the code will certainly have to cover the four areas as outlined in the AVMSD 28b in order to transpose the directive; and below we provide research-based evidence for covering cyberbullying and online harassment, non-consensual sharing of intimate images, domestic violence, content promoting self-harm and suicide and content promoting and encouraging behaviour that characterises an eating disorder, should all be covered under one or more online safety codes. There is also no denying that the harms related to disinformation,³ online misogyny, toxic masculinity and the related harassment⁴ require urgent attention as well even though we do not provide evidence below on some of these in the interest of space, and as some of them go beyond the scope of our immediate expertise. Furthermore, there is an increasing risk of Artificial Intelligence-generated cyberbullying and harassment, especially via images and videos (deep-fakes), which merits particular attention⁵ (in conjunction with the EU AI Act⁶). We outline the rationale as per each risk, below:

- **Cyberbullying and online harassment:** ABC researchers found that it is difficult to assess the effectiveness of VSPS' moderation of cyberbullying on their platforms, even with the information that platforms themselves are currently providing in their transparency reports.⁷ Having in mind the recent research from Ireland and internationally, where children and young people identify cyberbullying and online harassment on social media platforms as a significant problem; and that reporting to these platforms does not always help solve the issue,⁸ we find that it would be important

² <https://onlinesafety.org.au/codes/> and <https://www.esafety.gov.au/sites/default/files/2021-09/eSafety%20Industry%20Codes%20Position%20Paper.pdf>

³ Culloty, E., & Suiter, J. (2021). Anti-immigration disinformation. In *The Routledge companion to media disinformation and populism* (pp. 221–230). Routledge; Siapera, E. (2022). Platform Governance and the “Infodemic”. *Javnost-The Public*, 29(2), 197–214;

⁴ Ging, D. (2019). Alphas, betas, and incels: Theorizing the masculinities of the manosphere. *Men and masculinities*, 22(4), 638–657; In the context of Andrew Tate, please see: <https://www.irishtimes.com/opinion/2023/01/24/why-influencers-like-andrew-tate-want-your-sons-attention/>

⁵ See here: <https://cyberbullying.org/generative-ai-as-a-vector-for-harassment-and-harm>

⁶ <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A52021PC0206>

⁷ Verma, K., Milosevic, T., Davis, B. (2022). *Examining the Effectiveness of Artificial Intelligence-based Cyberbullying Moderation on Online Platforms: Transparency Implications*. Selected Papers of #AoIR2022: The 23rd Annual Conference of the Association of Internet Researchers, Dublin, Ireland.

⁸ Milosevic, T., Verma, K., Carter, M., Vigil, S., Laffan, D., Davis, B., & O'Higgins Norman, J. (2023). Effectiveness of Artificial Intelligence-Based Cyberbullying Interventions From Youth Perspective. *Social Media+ Society*, 9(1), Article 20563051221147325; Milosevic, T., & Vladislavljevic, M. (2020). Norwegian children's perceptions of effectiveness of social media companies' cyberbullying policies: an exploratory study. *Journal of children and media*, 14(1), 74–90

that the codes address these risks in some form. Cyberbullying is often considered as interchangeable with harassment in companies' policies, although the research community has more specific definitions for cyberbullying. Our research with 12-17-year-olds in IE also found that children have various understandings of what cyberbullying and harassment are.⁹ **Therefore, the online safety code should, in our view, take into account the relevant definitions of the phenomena and how social media platforms operationalise them for moderation.** When designing the online safety code, the Commission should also consider how the provisions in the Harassment, Harmful Communications and Related Offences Act 2020 (Coco's Law) would complement the provisions in the code.

- **Self-harm, suicide and content related to promotion of eating disorders:** While working on one of our current research projects. Cilter,¹⁰ funded by the Disruptive Technologies Innovation Fund, we have been unable to find sufficient evidence from industry and academic research to understand and independently evaluate how effective the technologies aimed at automatic detection of these types of risks on social media platforms are. In light of the high-profile cases internationally where the availability of self-harm promoting content on social media was linked to tragic incidents of children dying by suicide,¹¹ **we find that it would be important to ensure that the technology employed by VSPS in this regard is effective at limiting availability of such content to minors.**
- **Non-consensual sharing of intimate images:** Recent ABC research has provided an overview of facets, prevalence and legislation on image-based sexual abuse in Ireland and internationally.¹² In 2021, Foody and colleagues¹³ conducted a study on the prevalence of the solicitation of sexual images, as well as sending unwanted sexual images among adolescents (15-18 years-old). Results showed that 52.5% had been asked to send a naked image of themselves, but it was unclear if this solicitation was unwanted or individuals felt pressured to send the image. Further, 44% had received an unwanted sexual image and 29.5% reported that this was a frequent occurrence. However, in regards to research on this specific area, there is a dearth of literature from an Irish context.¹⁴ **Therefore, considering the urgent calls to tackle image-based sexual abuse in Ireland, it is our view that the online safety code should include non-consensual sharing of intimate images and how social media platforms identify this for moderation purposes.** The Commission should consider how the Harassment, Harmful Communications and Related Offences Act 2020 (Coco's Law)

⁹<https://antibullyingcentre.ie/wp-content/uploads/2022/11/Co-Designing-AI-Based-Cyberbullying-Interventions-on-Social-Media.pdf>

¹⁰ <https://www.cilter.ie/>

¹¹ <https://www.nytimes.com/2022/10/01/business/instagram-suicide-ruling-britain.html>

¹² Andreasen, M. B., Mazzone, A., Foody, M., Milosevic, T., & O'Higgins Norman, J. (2022). *The Gendered Experiences of Image-based Sexual Abuse: State of the Research and Evidence-based Recommendations*. Retrieved from <https://antibullyingcentre.ie/wp-content/uploads/2022/02/DCU-Online-Abuse-Report.pdf>

¹³ Foody, M., Mazzone, A., Laffan, D. A., Loftsson, M., & O'Higgins Norman, J. (2021). "It's not just sexy pics": An investigation into sexting behaviour and behavioural problems in adolescents. *Computers in Human Behavior*, 117, Article 106662.

¹⁴ Finn, C. (2021, December 23). *Justice Minister says 'stark' findings shows one in 20 have had intimate images shared online*. *Thejournal.ie*. <https://www.thejournal.ie/helen-mcentee-cocos-law-5638704-Dec2021/>

would complement the provisions in the code, as non-consensual sharing of intimate images is an offence recognised under The Harassment, Harmful Communications and Related Offences Act 2020.

- **Intimate Partner Violence:** ABC researchers have noted that there is dearth of research on intimate partner violence (IPV), specifically coercive control, and technology/social media among young people. In 2021, Women’s Aid published findings on young people’s (aged 18-25) understanding of IPV, including coercive control.¹⁵ Using a nationally representative sample in Ireland, the findings demonstrate that 3 in 5 young people have experienced, or know someone who has experienced, IPV. The results found that a majority of young people felt a responsibility to intervene if they are concerned about a close friend who might be experiencing IPV. However, the report also highlighted that only 16% of the sample believed it was easy to spot signs of IPV in a friend’s relationship. To our knowledge, this is the only piece of recent research investigating young people’s understanding of IPV, including coercive control, in Ireland.¹⁶ **Thus, the findings highlight that the importance of supporting young people to have their own agency is a crucial step to empowerment within their intimate relationships. The above research does not highlight how technology plays a role in IPV (including coercive control), but considering how entrenched it is within young people’s lives, it may be used as a means to perpetrate acts of IPV.**

Question 2: What types of online harms do you think should attract the most stringent risk mitigation measures by VSPS? How could we evaluate the impact of different types of harms e.g. severity, speed at which harm may be caused? Is there a way of classifying harmful content that you consider it would be useful for us to use?

Regarding the impact on children, from our perspective, it is very difficult to understand what is meant by “the most stringent risk mitigation measures” and by “severity of harm.” Is an instance of harm more severe if it impacts more users; or is it enough for it to have a strong negative impact on one user to be considered severe? For example, a cyberbullying incident, whereby a child is mocked, may have only a slight negative impact on a greater number of children who witness it, yet a severe negative impact on the one child who is targeted. Companies could be requested to ensure that fewer users see such content (i.e., to regulate based on ensuring that as few users as possible view a piece of harmful online content, see

¹⁵ Women’s Aid. (25th November, 2021). *Tackling Intimate Relationships Abuse among 18-25 year olds: Considerations for a Peer Supported Approach*. Retrieved from https://www.womensaid.ie/assets/files/pdf/tackling_intimate_relationship_abuse_among_18-25s_-_considerations_for_a_peer_supported_approach.pdf

¹⁶ Researchers in Northern Ireland investigated young people’s understanding of coercive control, please see the following article for further information: Lagdon, S., Klencakova, L., Schubotz, D., Shannon, C., Tully, M. A., Armour, C., & Jordan, J. A. (2023). Young People’s Understanding of Coercive Control in Northern Ireland. *Journal of Child & Adolescent Trauma*, 1–9.

here¹⁷). Under such circumstances, the bullying content that was viewed by fewer users may be underprioritised by the platform, yet it may still cause severe harm to the bullied child that is in a particularly vulnerable situation. Therefore, from our perspective, it is necessary to define severity in order to understand what is meant by stringent measures and how to design these measures.

Generally, there is an understanding that the harms that fall under illegal content such as child sexual abuse material or imagery (CSAM/CSAI) are particularly severe because of the effect on children involved. If the proposed EU Regulation on Combating Child Sexual Abuse (COM/2022/209)¹⁸ is adopted, it would also be essential not to replicate but to complement some of its requirements in the online safety codes. Similarly, the relevant online safety codes for CSAM/CSAI should complement the global multistakeholder strategic response approach, and likewise the model national response, as laid out by WeProtect Global Alliance.¹⁹

The Better Internet for Kids report²⁰ provides a helpful framework for classifying online risks and subsequent harms. Based on the EU Kids Online research framework,²¹ risk can be distinguished from harm, which can be particularly informative when attempting to define severity. It also classifies risks into content, conduct, contact and contract-related risks (4 Cs). The results of the UK Online Safety Data Initiative's Taxonomy project whereby they classified online harms, could also be instructive in this respect,²² and there is further research that has attempted to classify severity of harmful online content.²³

As per the The Australian Online Safety Commissioner position paper, an outcomes-based codes model is preferable to cultivate multi-stakeholder accountability, transparency and progress overall.²⁴

Question 3: Do you have reports, academic studies or other relevant independent research that would support your views? If you do, please share them with us with links to relevant reports, studies or research.

Recent ABC research indicates that minorities and girls in Ireland are especially negatively impacted by online harms. For example, research with young people about sexual and gender-based harassment during the Covid-19 pandemic showed that girls in Ireland experienced more online harms than boys; and LGBTQ+ students experienced more online harms than

¹⁷https://about.fb.com/wp-content/uploads/2020/02/Charting-A-Way-Forward_Online-Content-Regulation-White-Paper-1.pdf

¹⁸ <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=COM%3A2022%3A209%3AFIN>

¹⁹ <https://www.weprotect.org/response/>

²⁰ Stoilova, M., Rahali, M. & Livingstone, S (2023) *Classifying and responding to online risk to children: Good practice guide*. London: Insafe helplines and the London School of Economics and Political Science (LSE).

²¹ <https://www.lse.ac.uk/media-and-communications/research/research-projects/eu-kids-online>

²² <https://onlinesafetydata.blog.gov.uk/>

²³ See e.g.: Scheuerman, M. K., Jiang, J. A., Fiesler, C., & Brubaker, J. R. (2021). A framework of severity for harmful content online. *Proceedings of the ACM on Human-Computer Interaction*, 5(CSCW2), 1–33.

²⁴ <https://www.esafety.gov.au/sites/default/files/2021-09/eSafety%20Industry%20Codes%20Position%20Paper.pdf>

heterosexual students.²⁵ Furthermore, the same study found that girls experienced more sexual harassment than boys since the Covid-19 pandemic: twice as many girls (33.3%) as boys (17.4%) received unwanted sexual photos from friends, adult strangers or people they did not otherwise know offline. ABC research with the adult population in Ireland found that just under a half of respondents in Ireland have at one point experienced some form of online hate because of their personal identity or beliefs (such as race, ethnicity, gender, nationality, sexual orientation, religion, age, disability, etc.); those between the age of 18 and 25 were significantly more likely to experience online hate compared to older age cohorts (35 years old and up). Sexual minorities, people with disabilities and those belonging to the faith of Islam were particularly more likely to experience online hate.²⁶ In the case of post-primary students, ABC research found that 45.3% reported that they had witnessed cyberbullying at least once over the last few months, more frequently in the form of name calling, mockery or insults and occurs most often on social media.²⁷ The perceived reasons behind the victimisation identified in this report proved to be quite diverse, but those relates to sexual orientation and being overweight were the two most reported (by 22.9% and 21.9% of witnesses respectively).

In reference to evidence about the types of harmful online content that children in Ireland find to be the most upsetting, you can consult the NACOS study²⁸ which indicates that 22% of all 9-17 year olds who were bothered by something online said that bullying was most upsetting to young people their age; 19% said this was the case with inappropriate or disturbing images/photos, 9% said this was animal cruelty and 3% said it was material related to extreme dieting (NACOS report, p. 40).²⁹ As for seeing sexual images (which could refer to pornography, but also sexting-related images), 5% of all children who'd seen such images said they felt "very upset" after seeing them, whereas 10% said they were "a little upset," and 47% felt neutral about it (NACOS report, p. 59). This is *not* to suggest that exposure to pornography may not have long-term negative consequences,³⁰ but just to underscore that defining risk vs. harm, and also consequently how severity is defined, is important.

For a broader and more comprehensive picture about the prevalence of online risks and harms for children in Europe, you can consult the EU Kids Online's most recent research report from

²⁵ Ging, D. & Castellini da Silva, R. (2022). *Young people's Experiences of Sexual and Gender-based Harassment during Covid-19 Pandemic in Ireland: Incidence, Intervention and Recommendations*. Available at <https://antibullyingcentre.ie/wp-content/uploads/2022/10/Young-Peoples-Experiences.pdf>. The findings are based on a sample of 185 Transition Year students (15-17 years of age) in 2 co-educational Dublin schools.

²⁶ Andreasen, M. B., & McCashin, D. (2023). *Understanding Adult Experiences of Online Hate in Ireland: An Exploratory Survey*. DCU Anti-Bullying Centre. Available at <https://antibullyingcentre.ie/wp-content/uploads/2023/02/Understanding-adult-experiences-of-online-hate-in-ireland-2023-Final.pdf>

²⁷ Feijóo, S., Sargioti, A., Sciacca, B., & McGarrigle, J. (2023). *Bystander Behaviour Online Among Young People in Ireland*. DCU Anti-Bullying Centre, Dublin City University. Available at <https://antibullyingcentre.ie/wp-content/uploads/2023/05/Bystander-Behaviour-Online-Report.pdf>

²⁸ Nationally representative sample of 9-17-year-old Internet-using children

²⁹ <https://www.gov.ie/en/publication/1f19b-report-of-a-national-survey-of-children-their-parents-and-adults-regarding-online-safety/>

³⁰ Owens, E. W., Behun, R. J., Manning, J. C., & Reid, R. C. (2012). The impact of internet pornography on adolescents: A review of the research. *Sexual Addiction & Compulsivity*, 19(1-2), 99–122.

19 European countries (nationally representative samples of 9-16-year-old Internet-using children).³¹

Question 4: What approach do you think we should take to the level of detail in the Code? What role could non-binding guidance play in supplementing the Code?

A very detailed, prescriptive code (Option 1) may fail to take differences in technological affordances of various platforms into account and create unintended chilling effects (e.g., incentivise companies to remove even legitimate content to ensure compliance). Option 2, a very high-level set of guidelines might, on the other hand, allow companies to evade adequate scrutiny. Option 3—Imposing high level obligations with supplemental guidelines might be the optimal approach. It is important for us to emphasise that we find that it is often insufficient for companies to simply provide metrics about their content moderation. For example, in their Transparency Reports, some companies are already providing metrics about their rates of proactive (Artificial Intelligence or AI-based) harmful content removal,³² sometimes with impressive rates of removal. However, in non-industry research, young users complain that they still see and are bothered by such content on social media platforms, as was the case in our recent report into the prevalence of online hate in Ireland.³³ In our view, it is important for the Online Safety Commissioner’s office to examine the effectiveness of content moderation by conducting evaluation research with children and adults alike; and by inquiring with them directly about their experiences on the said platforms. Research has previously found that young people highlight the lack of platforms’ responsiveness to their reporting of harmful online content, and it continues to dominate their concerns.³⁴

Question 5: What do you think would be the most effective structure for the Code? What are the most important factors we should consider when we decide how to structure the Code?

As mentioned, The Australian Online Safety Commissioner's Guidelines for the industry association’s development of their own codes might be helpful here as well. They broadly designed the objectives; specific outcomes under each objective; and compliance criteria per outcome.³⁵ We suggest the Commission might adopt similar best practices when formulating something similar for each category or several categories of harmful online content, and we

³¹ Smahel, D., Machackova, H., Mascheroni, G., Dedkova, L., Staksrud, E., Ólafsson, K., Livingstone, S., & Hasebrink, U. (2020). *EU Kids Online 2020: Survey results from 19 countries*.

³² Milosevic, T., Van Royen, K., & Davis, B. (2022). Artificial intelligence to address cyberbullying, harassment and abuse: New directions in the midst of complexity. *International journal of bullying prevention*, 4(1), 1–5.

³³ Andreasen, M. B., & McCashin, D. (2023). *Understanding Adult Experiences of Online Hate in Ireland: An Exploratory Survey*.

³⁴ Milosevic, T., & Vladislavljevic, M. (2020). Norwegian children’s perceptions of effectiveness of social media companies’ cyberbullying policies: an exploratory study. *Journal of children and media*, 14(1), 74–90; Also, most recent findings from Ireland from yet unpublished PhD research from ABC and ADAPT Science Foundation Ireland researcher Kanishk Verma, on a small convenience sample (N= 151) of 18-21-year-olds in Ireland, 59% disagreed with the statement that social media platforms are responsive when mean or hurtful content is reported to them.

³⁵ <https://onlinesafety.org.au/codes/>

reiterate our position on the importance of distinguishing between online risk and harm as previously stated in reply to Question 2.

Question 6: How should we design the Code to minimise the potential for conflict and maximise the potential for synergies in how platforms comply with it and the DSA? How should the Code take account of the Digital Services Act (“DSA”)?

The code could be based on the high-level requirements outlined in the DSA and take these as the minimum requirements for the platforms and develop more detailed guidelines for effective compliance. In our understanding, DSA contains provisions for voluntary codes of conduct only (Recital 103 p. 28, Articles 45, 46), whereas the codes under the OSMR are binding for the designated platforms. Hence, we are not sure how these two provisions might conflict.

Question 7: To what extent, if at all, should the Code require VSPS providers to take measures to address content connected to video content?

Cyberbullying takes place not only in videos but often takes place in conjunction with comments, videos, and images, rather than being limited to videos alone.³⁶ In fact, certain video reels may not be harmful on their own, but it is the related comments and text that constitute the bullying behaviour. Therefore, it is crucial to take a holistic approach when addressing platform content, going beyond merely regulating video content. For example, based on the most recent findings we have³⁷— 59% of young people who saw mean or hurtful content saw it in comments; 32% saw it in video reels with text captions; and 24% in images with text captions; only 16% said they saw it in video reels only and another 9% in images only.

These findings indicate that a considerable portion of cyberbullying occurs outside of the video content itself, and addressing these issues requires a comprehensive strategy that includes comments, captions, and images as well. Moreover, textual captions provide an additional layer of context and interpretation to the visual content, which can be misused to spread harmful messages, harassment or offensive language. VSPS providers should be encouraged to implement stringent moderation systems that analyse the content of both videos and their associated textual captions and comments. Furthermore, VSPS providers should also take initiative to educate their users about the importance of thoughtful and considerate captioning.

³⁶ Milosevic, T., Verma, K., Carter, M., Vigil, S., Laffan, D., Davis, B., & O’Higgins Norman, J. (2023). Effectiveness of Artificial Intelligence–Based Cyberbullying Interventions From Youth Perspective. *Social Media + Society*, 9(1), Article 20563051221147325; <https://act-agi.github.io/>

³⁷ From a small, convenience sample of young people from Ireland (N=151, age 18-21) conducted by PhD researcher Mr. Kanishk Verma.

All of the above should take into account the added complexities of live-streaming audio-visual content which, according to the most recent evidence synthesis, may necessitate its own legal framework.³⁸

In recent months, the landscape of audio-visual creation has been revolutionised by Generative Artificial Intelligence (AI) tools,³⁹ which have made content creation accessible to all age demographics. Simultaneously, the emergence of deepfake technologies, capable of manipulating real videos to create fabricated yet convincing videos, has equipped video content creators as a new avenue for audio-visual creation. While the effects of such fabricated audiovisual content on propagation of cyberbullying and other online are yet to be fully grasped, there have been reports that suggest this is of growing concern.⁴⁰ To that effect, VSPs providers should incorporate distinct markers or labels, as outlined in for videos generated using AI. By doing so, these platforms can actively contribute to preventing the potential misuse of AI-generated content and fostering a safer digital environment for all users.

Question 8: How should we ask VSPS providers to introduce a feature that allows users to declare when videos contain advertising or other type of commercial communications? Should the Code include specific requirements about the form in which the declaration should take? What current examples are there that you regard as best practice?

Given the most recent evidence from Ofcom in the UK which found that most users (especially children) struggle to fully understand commercial communications.⁴¹ Coupled with emerging evidence on the use of dark patterns of design within industry to engage and sometimes manipulate users,⁴² the Code should ensure that no such unethical practices can continue. With this evidence base in mind, all proposed features relating to advertising declarations and all commercial communications should be built with the following principles: informed consent, age appropriateness, evidence-based engagement strategies, and neurodiversity. Furthermore, it is advisable to consider how the Code could engage audits of such features (for example, see sludge audits as proposed in footnote 34). All these considerations should be applicable to the full range of advertising and commercial content, including the often blurred lines between advertising and brand ambassadorship, product placements, and the role of influencer culture.

³⁸ Drejer, C., Riegler, M. A., Halvorsen, P., Johnson, M. S., & Baugerud, G. A. (2023). Livestreaming technology and online child sexual exploitation and abuse: a scoping review. *Trauma, Violence, & Abuse*, Article 15248380221147564.

³⁹ <https://phenaki.video/> ; <https://arxiv.org/abs/2304.08551>

⁴⁰ Abah, A. L., & Sanders, A. K. (2022). Obscenity, Nonconsensual Pornography, and Cyberbullying. In *Social Media and the Law* (pp. 150-169). Routledge.

⁴¹ <https://www.ofcom.org.uk/news-centre/2023/video-sharing-platforms-rules-kids#:~:text=Ofcom's%20study%20finds%20that%20the,for%20many%20users%2C%20including%20children>

⁴² Mills, S., Whittle, R., Ahmed, R., Walsh, T., & Wessel, M. (2023). Dark patterns and sludge audits: an integrated approach. *Behavioural Public Policy*, 1–27.

Question 9: How should we ask VSPS providers to introduce and design a flagging mechanism in the Code? How can we ensure that VSPS providers introduce the mechanism in a user-friendly and transparent way? How should we ask VSP Providers to report the decisions they've made on content after it has been flagged? To what extent should we align the Code with similar provisions on flagging in the DSA?

It might be helpful to consult the wording used in the Australian Online Safety Code for Social Media Platforms for regulating class 1A and 1B content in that respect (Schedule 1 – Social Media Services Online Safety Code [Class 1A and Class 1B Material], 31 March 2023, p. 17).⁴³ In our view, it would be important to require the companies to be able to demonstrate that they have engaged children in the design of their flagging tools and that children find them easy to understand and navigate. Likewise, it would be important that companies demonstrate that information they give to children as to company's decisions about the reported content is easy for children to understand and does not operationalise dark patterns (defined as: 'tricks used in websites and apps that make you do things you didn't mean to, like buying or signing up for something').⁴⁴

Question 10: What requirements should the Code include about age verification and age assurance? What sort of content should be shown by default to users who are logged out or in private browsing mode and whose age cannot be verified or assured? What evidence is there about the effectiveness of age estimation techniques? What current practices do you regard as best practice? Where accounts are not age verified should default privacy settings be used, should content default to universal content and should contact by others be more limited?

In light of significant underage use of social media in Ireland (see e.g. the NACOS report, 2021, p.24), we find it important that the code places transparency requirements on companies to disclose how specifically they conduct age verification and assurance; and to provide information on effectiveness of this process. Despite significant advancements in terms of industry understanding of age assurance and age verification, as well as in terms of technologies that are being used for such purposes.⁴⁵ It still appears to be difficult to understand how specifically VSPS engages in age verification and age assurance, how effective it is, and how such processes adhere to the rights of the child and GDPR. Indeed, it is noteworthy that in Ireland and the UK, there have been significant fines for data protection breaches by dominant social media companies, in particular for breaches of children's privacy.⁴⁶

⁴³ <https://onlinesafety.org.au/codes/>

⁴⁴ Mathur, A., Acar, G., Friedman, M. J., Lucherini, E., Mayer, J., Chetty, M., & Narayanan, A. (2019). Dark patterns at scale: Findings from a crawl of 11K shopping websites. *Proceedings of the ACM on Human-Computer Interaction*, 3(CSCW), 1–32.

⁴⁵ <https://www.biometricupdate.com/202212/euconsent-reports-improved-age-verification-tests-yoti-gets-another-partner>; and <https://www.yoti.com/business/age-verification/>

⁴⁶ See recent coverage of these breaches here: <https://www.theguardian.com/technology/2023/aug/04/tiktok-to-be-fined-for-breaching-childrens-privacy-in-eu>

We understand the privacy and freedom of expression-related concerns when mandating document-based age-verification from all users, and we would not support those. Advancements in biometric age verification in terms of data minimisation and minimising potential for privacy infringements appear promising, but we have not conducted research ourselves in this domain to be able to ascertain this claim.

Transparency in terms of effectiveness of age-assurance and verification procedures appears to be lacking. Our researchers recently informally asked several VLOPs to explain the process of how they verify the age of, for example, a 9-year-old who attempts to open an account on their services, and we were not able to receive clear answers. Do all children undergo biometric verification? Is parental consent sought in all instances where the child declares they are under the digital age of consent as mandated in the given location, as per Article 8 of the General Data Protection Regulation, when prompted to insert their age? If the child is allowed on the platform but they are under the digital age of consent, and if it is subsequently determined that the child was underage at the time of sign-up, what happens to the child's personal data if the company failed to seek parental consent for its processing? We were not able to receive answers to these questions, which appear to be very simple compliance-related questions.

Most importantly, if companies deny that they have underage users on their platforms, they are not obliged to create policies that are age-appropriate for them. Legislation that incentivises companies to assert ignorance of underage use on their platforms in order to avoid liability, also disincentivises companies to innovate for underage users.⁴⁷

Furthermore, we find it important to articulate a clear policy that is understandable to the public that age limits in companies' policies (being 13) are a by-product of privacy legislation (Children's Online Privacy Protection Act and The GDPR). These limits are not there for safety reasons but are misleadingly utilised to such effect; and that as long as companies proclaim not to have the actual knowledge of underage users on their platforms, they are not in breach of such law. At the same time, maturity differs from child to child, not all children become magically mature at age 14; and therefore age-based cut-offs can be inherently problematic for policy design.

Question 11: What requirements should the Code have in relation to content rating? What do you consider to be current best practice? What experiences have you had using content rating systems on platforms and do you think they have been effective? What steps could we ask VSPS to take to ensure content is rated accurately by users?

Content rating systems employed by major VSPs providers, such as YouTube, Twitch, TikTok, identify and categorise content to provide platform users with a clear understanding of its

⁴⁷ Boyd, D. (2015, December 18). What if social media becomes 16-plus? New battles concerning age of consent emerge in Europe. *The Medium*. Retrieved from <https://medium.com/bright/what-if-social-media-becomes-16-plus-866557878f7#skvnifxhd>; Boyd, D., Hargittai, E., Schultz, J., & Palfrey, J. (2011). Why parents help their children lie to Facebook: Unintended consequences of Children's Online Privacy Protection Act (COPPA). *First Monday*, 16(11). Retrieved from <http://firstmonday.org/ojs/index.php/fm/article/view/3850/3075>.; Milosevic, T. (2018). *Protecting children online?: Cyberbullying policies of social media companies*. The MIT Press.

maturity level. TikTok employs a "Content Levels" system,⁴⁸ whereas Twitch employs "Content Classification" system,⁴⁹ and YouTube utilises its own content rating mechanism.⁵⁰ Despite their different names, these systems generally adhere to similar guidelines and practices for content classification. Content is typically rated based on factors such as strong language (encompassing more than just vulgarity and profanity), nudity, mature-rated games, sexual themes, drug use, and violent/disturbing content, among other criteria. This structured approach helps viewers understand the nature of the content they are about to engage with, allowing for more informed choices about what they consume.

Recent months have witnessed the proliferation of Generative Artificial Intelligence (AI) tools that enable users of all age groups to create synthetic content.⁵¹ Additionally, the rise of deepfake technology, which generates manipulated videos from real footage, poses both positive and negative implications.⁵² These rapidly evolving technological landscapes have introduced new challenges to content rating systems. To stay current with upcoming technologies in video generation, VSPS providers could add flag videos that are created using DeepFake or Generative AI. For example, videos employing DeepFake to portray another actor discussed in the interview⁵³ showcasing the potential of DeepFake, could be accompanied by labels or flags indicating it is generated using AI. This approach will enhance transparency and innovation awareness. In this dynamic environment, content rating systems must adapt to these emerging complexities to remain effective and relevant.

Potentially one of the biggest concerns of both DeepFake and Gen AI is its potential to be used in cyberbullying and other forms of online harm if left unregulated through this Code. Already there have been reports that indicate a surge in the use of such technologies to inflict repeated types of online harm on others.⁵⁴ To that effect, through this Code, VSPS providers should be required to include flags to labels for such AI-generated video in their content rating system.

Additionally, the Pan-European Game Information (PEGI) age rating system was established in 2003 to provide age classifications for video games in 38 European countries, becoming one of the most prominent international content ratings for age appropriateness of electronic content.⁵⁵ However, it has been shown that the PEGI system does not provide the intended guidance to consumers to ascertain if a particular product is adequate for a child, but it is necessary to raise awareness about the system and ensure that the rating is clear enough to the target audience and they know how to use it. Furthermore, the PEGI system is a video game industry self-audit, raising doubts about the system integrity and its appropriateness to protect

⁴⁸ <https://newsroom.tiktok.com/en-us/more-ways-for-our-community-to-enjoy-what-they-love>

⁴⁹ <https://blog.twitch.tv/en/2023/06/20/introducing-content-classification-labels/>

⁵⁰ <https://support.google.com/youtube/answer/4601348?hl=en#:~:text=A%20YouTube%20content%20rating%20contains,separate%20the%20values%20with%20spaces.>

⁵¹ <https://phenaki.video/> ; <https://arxiv.org/abs/2304.08551>

⁵² <https://arxiv.org/abs/2105.00192>

⁵³ <https://www.youtube.com/watch?v=VWrhRBb-1Ig>

⁵⁴ Abah, A. L., & Sanders, A. K. (2022). Obscenity, Nonconsensual Pornography, and Cyberbullying. In *Social Media and the Law* (pp. 150-169). Routledge; Mullen, M. (2022). A new reality: deepfake technology and the world around us. *Mitchell Hamline Law Review*, 48(1), Article 5.

⁵⁵ <https://pegi.info/>

the wellbeing of children without being biased towards the interests of the industry, recommending instead a system managed by another body even if it collaborates with the industry in order to provide the age recommendations.⁵⁶ Finally, the content rating needs to have a pedagogical point of view to avoid contradictions, while at the same time taking into account the rights of the children. In this sense, PEGI and Entertainment Software Rating Board (ESRB)⁵⁷ content rating mechanisms have been inconsistent in labelling micro transactional features in multimedia such as games (e.g. Loot Boxes). Inconsistencies that favour developers and industry more broadly in this area mean that parents and children cannot entirely rely on ratings alone.⁵⁸ Therefore, other complementary parental strategies whereby parents can assess the content in advance should be advised, such as viewing movie clips and trailers themselves before allowing children access to a particular content, or reading the reviews online and analysis from trusted journalists and media.

Question 12: What requirements should the Code have in relation to parental control features? How can we ensure that VSPS providers introduce the mechanism in a user-friendly and transparent way? Can you point to any existing example of best practice in this area? Should parental controls be ‘turned-on’ by default for accounts of minors or where age is not verified?

Parental control technologies are not universally effective and they need to be tailored to the evolving capacities of the child. It is important to emphasise that restricting access to technology or certain features does not necessarily reduce the risk of harm. Use of parental controls can also have negative effects: it can render certain behaviours and technologies more appealing, resulting in poorer decision-making and less resilience on behalf of the child.⁵⁹ Monitoring and surveillance can also disrupt parent-child relationship and trust, and have a negative impact on a child's right to privacy.⁶⁰ An effective parental control tool will likely vary by technological affordances of the platform; and by the child's age. Please also consider disadvantaged children who turn to social media platforms because they are suffering abuse at home; or because they may not encounter emotional support that they need, which they satisfy through online relationships or advice from friends or professionals online, often via social media platforms. Mandating parental control can be counterproductive in those cases. Finally, it must be acknowledged that research has found that a positive approach to supervision seems

⁵⁶ Felini, D. (2015). Beyond today's video game rating systems: A critical approach to PEGI and ESRB, and proposed improvements. *Games and Culture*, 10(1), 106–122.

⁵⁷ <https://www.esrb.org/>

⁵⁸ Xiao, L. Y. (2023). Beneath the label: unsatisfactory compliance with ESRB, PEGI and IARC industry self-regulation requiring loot box presence warning labels by video game companies. *Royal Society Open Science*, 10(3), Article 230270.

⁵⁹ Smirnova, S., Livingstone, S., & Stoilova, M. (2021). *Understanding of user needs and problems: A rapid evidence review of age assurance and parental controls*. Retrieved from https://eprints.lse.ac.uk/112559/1/Stoilova_understanding_of_user_needs_and_problems_published.pdf

⁶⁰ Zaman, B., & Nouwen, M. (2016). Parental controls: advice for parents, researchers and industry. *EU Kids Online*, 1-9; Ghosh, A. K., Badillo-Urquiola, K., Guha, S., LaViola Jr, J. J., & Wisniewski, P. J. (2018). Safety vs. surveillance: what children have to say about mobile apps for parental control. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (pp. 1–14).

to be more efficient than a restrictive one.⁶¹ It has been suggested that there are four ways in which parents can influence the safe and responsible use of the Internet of their children: through “active mediation”, which are conversations between family and children to foster children's understanding and critical analysis of Internet content and usage, including norms of adequate use; through “co-use” or shared use of the Internet between family and children, with the main intent of establishing the parents as the role model for good use; through “restrictive mediation”, which consists of limiting time, activities and content either by a set of rules or using specific software; or “supervision”, with the family monitoring children's use of the Internet, whether overtly or covertly.⁶² We do not have any specific examples to recommend, other than that any guidelines for parental controls technologies in the code should be at a very high level – as guidelines that companies can make available to parents or caregivers. We would advise caution regarding any provisions in the code as to having them turned on by default for the reasons mentioned above.

Question 13: What requirements should the Code contain to ensure that VSPS provide for effective media literacy measures and tools?

In our view, the code should request that VSPS provide clear and age-appropriate guidelines about its enforcement mechanisms such as blocking, reporting, and muting with step-by-step guidelines as to how to engage in this on the platform itself. It would also be important to adapt the language of such education to younger children, as well those who are below the platform's stipulated age of access as per Terms of Service because of under-age use.

Some platforms already have educational content (typically developed by experts or advocacy organisation representatives and sometimes academic institutions) dedicated to online safety or digital citizenship advice for parents and children, (e.g., Family Centres on Facebook/Meta, Instagram and Snapchat).⁶³ However, it is not entirely clear to what extent children are aware of these resources, whether they use them and find them engaging and helpful. For example, in a recent study in Norway, we found that a large portion of children surveyed were not aware of Safety Centres.⁶⁴ Hence, we strongly believe that it would be important to ensure the following: (1) That the advice in these educational resources is based on research evidence, i.e. that the effectiveness of online safety and media literacy advice is independently evaluated with children and parents in terms of accessibility and awareness and willingness to listen to such advice; (2) It is important that this advice does not serve merely as a branding tool or a box-ticking exercise for companies to showcase that they are doing something to assuage the harmful consequences of risks children can experience on their platforms. **We therefore**

⁶¹ Feijóo, S. (2022). *Problematic Internet Use and online risk behaviors. An analysis from the gender perspective*. Universidade de Santiago de Compostela. <https://minerva.usc.es/xmlui/handle/10347/28872>

⁶² Nikken, P., & Jansz, J., (2013). Developing scales to measure parental mediation of young children's internet use. *Learning, Media and Technology*, 39(2), 250–266.

⁶³ See e.g. [https://about.meta.com/actions/safety/audiences/childsafety/;](https://about.meta.com/actions/safety/audiences/childsafety/)
<https://help.instagram.com/454886756318459>

and <https://help.snapchat.com/hc/en-us/articles/7121384944788-What-is-Family-Center->

⁶⁴ Milosevic, T., & Vladislavljevic, M. (2020). Norwegian children's perceptions of effectiveness of social media companies' cyberbullying policies: an exploratory study. *Journal of children and media*, 14(1), 74–90.

recommend that the effectiveness of online safety and digital citizenship⁶⁵ advice provided be periodically evaluated by the Commission.

We would also recommend that the Commission consider stipulating a media literacy levy for the VSPS or at least Very Large Online Platforms (as per DSA), if this is lawfully possible, whereby the companies would be obliged to invest in evidence-based education focused on resilience building, digital skills or citizenship and wellbeing of minors. The levy could be determined by the Commission and distributed to relevant organisations via an open competition (e.g. a tender for bidding media literacy organisations which could be advocacy organisations, academic institutions or other organisations with adequate capacity for delivery of such education). We think it would be important that the Commission administer this process, rather than having the companies self-regulate by deciding on such education on their own. The education needs to be evidence-based⁶⁶ and decided upon by experts in the field, rather than companies themselves. If companies are left to decide on how they will implement such education and which organisations they will commission to implement it, then such provisions could place these educational institutions into a dependent position *vis-a-vis* VSPS, which can undermine the effectiveness of such education.

In a similar vein, we believe that VSPS could be levied to provide financial support for organisations that administer helpline services. As previously documented⁶⁷ issues encountered on VSPS are largely outsourced to helplines and trusted flaggers that are obliged to conduct such work without being able to charge VSPS for their services. Often, they are advocacy organisations that rely on state, EU or private forms of funding. If VSPS should provide funding for them voluntarily (without being required to do so by law) such arrangement can place helplines in a disadvantageous, dependent position *vis-a-vis* VSPS. This is why it would be advisable for the Commission to consider if such a levy can be formalised through the Commission and instituted in a way that helplines can benefit by applying for such levy-originating funding directly from the Commission.

Question 14: How should we ask VSPS providers to address online harms in their terms and conditions in the Code, including the harms addressed under Article 28b? How should key aspects of terms and conditions be brought to users' attention? What examples are there of best practice in relation to terms and conditions including content moderation policies and guidelines?

It would be important, in our view, to go beyond the requirements of transposing AVMSD by regulating not only video but text-based content (please see our points on this matter answered

⁶⁵ *Digital citizenship* is a concept that needs to be precisely defined and measured, rather than used as a catch-all phrase, in a tokenistic fashion, see: Jones, L. M., & Mitchell, K. J. (2016). Defining and measuring youth digital citizenship. *New media & society*, 18(9), 2063–2079. For other work on digital citizenship see Council of Europe's report, for example: <https://ec.europa.eu/newsroom/just/items/672450/en>

⁶⁶ Finkelhor, D., Walsh, K., Jones, L., Mitchell, K., & Collier, A. (2021). Youth internet safety education: Aligning programs with the evidence base. *Trauma, violence, & abuse*, 22(5), 1233–1247.

⁶⁷ Milosevic, T. (2018). *Protecting children online?: Cyberbullying policies of social media companies*. The MIT Press.

in other questions above). Many VSPS and especially VLOPs stipulate in their Terms of Service (ToS) and Community Guidelines/Standards that the types of content specified in AVMSD Article 28 (b) are not allowed on their platforms. Some platforms provide more detailed information than others, and Meta's community guidelines might provide a good example of a fairly elaborate explanation as to what is considered to be bullying and harassment on Facebook, for instance (examples of such behaviours are listed and it is stated that they are to be repeated⁶⁸ for the platform to take action).⁶⁹

In order to be able to hold platforms to account for bullying and harassment cases that they fail to remove from their platforms, it is important to understand how they define bullying and harassment and other potentially harmful content and behaviours. Previously, VLOPs would say that they cannot publicly reveal their internal definitions of harmful content and consequently operational instructions that they provide to their moderators because in doing so, they would risk abuse from "bad actors" who would now know how to act in order to circumvent the policy.⁷⁰ At the same time, previously leaked moderation guidelines revealed serious omissions in how platforms regulated harmful online content.⁷¹ Therefore, if the Commission chooses not to request from companies to reveal/publish operational moderator guidelines in the Community Guidelines, it would still be important for the Commission to be able to request access to such internal documents from companies, at least for auditing purposes.

Otherwise, we are of the opinion that the Commission should request in codes that ToS and/or Community Guidelines/Standards provide examples of banned harmful content/behaviours; as well as provide age-appropriate explanations as to how the platform decides as to what constitutes a violation and what does not constitute a violation. Age-appropriate ToS/Community Guidelines/Standards content can include videos for younger children and not merely text.

Question 15: How should we ask VSPS providers to address content moderation in the Code? Are there any current practices which you consider to be best practice? How should we address automated content detection and moderation in the Code?

For a high-level moderation requirement that should give significant flexibility to platforms to comply with moderation requirements, please consult the Social Media Services Online Safety Code from the Online Safety Commissioner of Australia which applies to illegal content.⁷² At the very minimum, the code should require VSPS to provide robust moderation capacity that can effectively respond to user complaints and remove illegal and harmful content that is

⁶⁸ For bullying, specifically, it would be important to emphasise that one-time hurtful actions that can be shared and replicated and seen by a wider audience should also be considered as bullying, and that repetition itself is a problematic concept because one-off acts can be equally hurtful.

⁶⁹ <https://transparency.fb.com/policies/community-standards/bullying-harassment/>

⁷⁰ Ibid. footnote 57.

⁷¹ <https://www.theguardian.com/news/2017/may/22/how-facebook-allows-users-to-post-footage-of-children-being-bullied>

⁷² <https://www.esafety.gov.au/industry/codes/register-online-industry-codes-standards>

against the policy; or otherwise sanction behaviours that contravene the companies' ToS. In line with DSA requirements, the codes should request that companies respond to user complaints in a timely manner, that they be informed about the process/steps and the outcome of the reporting process; and that there are measures of appeal. Prescribing a turnover time for complaints could lead to unintended effects of platforms prioritising content take-down of even legitimate content in order to ensure compliance; also different types of harmful content/behaviours may require different processing times. It is difficult to provide a recommendation as to how specific the codes should be in this regard.

From our perspective, it would be important that companies provide information about the percentage of their revenue that is invested in moderation and online safety; and to provide a more detailed account of such expenditure; to what extent moderation is facilitated by in-house vs. outsourced moderators; as well as how they are ensuring acceptable working conditions and support for moderators.⁷³

The aforementioned Australian Online Safety Code for Social Media Platforms makes it a requirement that companies provide automated proactive detection of illegal content such as child sexual abuse and extreme violence. Based on our previous research into automated (Artificial Intelligence-based) moderation of cyberbullying,⁷⁴ we think that the code should require that companies provide information on which automated and AI-based technologies they use to detect not just illegal but also harmful online content and behaviours and to provide information on effectiveness of such measures (more on that elsewhere in this submission); how they are executed (e.g. are direct messages monitored too and what the privacy implications are). Our recent research with young people from Ireland found that while they would welcome AI-based interventions into cyberbullying on social media, young people voiced concerns around privacy, transparency and freedom of expression of such automated monitoring and enforcement. They would like to know how AI-based proactive moderation is executed and they would also like to be able to opt-in and out of this process when it comes to cyberbullying detection.⁷⁵ For these reasons, we highly recommend that the codes incorporate young people's views on these matters; and mandate transparency with respect to AI use and other automated means of risk/harm detection; and that platforms demonstrate how their decisions on enforcement mechanisms incorporate young people's views.

⁷³ Roberts, S. T. (2019). Behind the screen. In Gillespie, T. (2018). *Custodians of the Internet: Platforms, content moderation, and the hidden decisions that shape social media*. Yale University Press.

⁷⁴ Ibid. footnotes 34 and 37; Verma, K., Milosevic, T., Cortis, K., & Davis, B. (2022, June). Benchmarking Language Models for Cyberbullying Identification and Classification from Social-media texts. In *Proceedings of the First Workshop on Language Technology and Resources for a Fair, Inclusive, and Safe Society within the 13th Language Resources and Evaluation Conference* (pp. 26-31); Verma, K., Davis, B., & Milosevic, T. (2022). Examining the Effectiveness of artificial intelligence-based cyberbullying moderation on online platforms: transparency implications. *AoIR Selected Papers of Internet Research*; Milosevic, T., Van Royen, K., & Davis, B. (2022). Artificial intelligence to address cyberbullying, harassment and abuse: New directions in the midst of complexity. *International journal of bullying prevention*, 4(1), 1-5.

⁷⁵ Milosevic, T., Verma, K., Carter, M., Vigil, S., Laffan, D., Davis, B., & O'Higgins Norman, J. (2023). Effectiveness of Artificial Intelligence-Based Cyberbullying Interventions From Youth Perspective. *Social Media+ Society*, 9(1), Article 20563051221147325.

Question 18: What approach do you think the Code should take to risk assessments and safety by design? Are there any examples you can point us towards which you consider to be best practice?

Again, it would be helpful to consult the Australian Online Safety Commissioner's Basic Online Safety Expectations (BOSE)⁷⁶ report which provides basic guidelines for platforms on how to ensure safety by design and the principles of safety by design. There is also the UK's government guidance on safety by design for online platforms.⁷⁷ We also recommend reviewing the Internet Commission's report on online platforms' maturity levels, which outlines the stages that companies go through in terms of building their online safety capacity. This document could be helpful in outlining the requirements.⁷⁸ Requiring the companies to conduct risk assessments is a provision for VLOPs in the DSA (Article 34) and we think it might be beneficial if the code contained a provision that would either require or recommend from VSPS in general to run periodic risk assessments which are adjusted to (which take into account) their scale and company capacity. VLOPs could be requested to provide annual reports on safety by design measures whereas other VSPS could be recommended to do the same. In our understanding, as per OSMR, the Commissioner is already entitled to request such information from designated companies. Lastly, it is essential that all risk assessments are grounded in an evidence-based approach that is domain specific (i.e., CSAM risk assessments need to be informed by relevant academic literature in addition to industry-recognised tools).

Question 19: How do you think that cooperation with other regulators and bodies can help us to implement the Code for VSPS?

If possible, the Global Online Safety Regulators' network could perhaps be asked to provide feedback on online safety codes and the European Board for Digital Services under DSA (once established); the same could be said of OfCom and the Data Protection Commission in IE as well as the UK's Information Commissioner's Office that developed the Age-appropriate design code. European Centre for Algorithmic Transparency⁷⁹ could be consulted regarding transparency and enforcement of proactive and overall automated and other AI-based harms regulations. It would also be important to coordinate with the entities responsible for developing the EU Regulation to Prevent and Combat Child Sexual Abuse (Comm 2022/209).⁸⁰ At a national level, it would be important to coordinate with the Ministry of Justice with respect to how the OSMR complements provisions in Coco's Law as well as harassment definitions therein (Harassment, Harmful Communications and Related Offences Act of 2020). It would also be important to coordinate with the Department of Education with respect to phone use in schools in particular and recent calls for bans; but also regarding developing

⁷⁶ <https://www.esafety.gov.au/industry/basic-online-safety-expectations> and subsequent report from the industry: <https://www.esafety.gov.au/sites/default/files/2022-12/BOSE%20transparency%20report%20Dec%202022.pdf>

⁷⁷ <https://www.gov.uk/government/collections/online-safety-guidance-if-you-own-or-manage-an-online-platform>

⁷⁸ <https://inetco.org/report>

⁷⁹ https://algorithmic-transparency.ec.europa.eu/about_en

⁸⁰ <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=COM%3A2022%3A209%3AFIN>

synchronised online safety, digital literacy and citizenship education that are based on research evidence and periodically evaluated for their effectiveness.

Question 22: What compliance monitoring and reporting arrangements should we include in the Code?

As indicated elsewhere, we find it important that compliance goes beyond requesting companies to provide long stats-heavy annual reports that contain rates of proactive removal of illegal and harmful content; but are not verified independently and that do not provide evaluation from the perspective of end-users, and children in particular. Such reports could turn into a box-ticking activity for VSPS and allow companies to cite high rates of proactive automated removal, while end users and children in particular still continue to encounter considerable harm on the platform. This is why we think it would be important to include provisions in the code that would allow the Commission to request independent bodies to conduct evaluation research with end-users and children in particular about the harms experienced on designated platforms.⁸¹ While such provisions already exist in the OSMR as regards to auditing company activity, we think it would be important to conduct such evaluation regularly, especially for VSPS that are popular with children, rather than merely on an ad-hoc basis.

Furthermore, following our findings that it is difficult for non-company-affiliated (independent) researchers to evaluate the effectiveness of companies' mechanisms for removal of cyberbullying (such as AI models),⁸² it might be advisable to synchronise the provisions in the Code with the DSA requirements from platforms to provide data access to independent researchers for evaluation (DSA Article 40, Data Access and Scrutiny). This would ensure that the data that companies provide are conducive to meaningful transparency,⁸³ transparency that allows us to understand how well the companies are performing from users' and children's perspective, rather than providing statistics as a box ticking exercise.⁸⁴

⁸¹ Milosevic, T., Verma, K., Carter, M., Vigil, S., Laffan, D., Davis, B., & O'Higgins Norman, J. (2023). Effectiveness of Artificial Intelligence-Based Cyberbullying Interventions From Youth Perspective. *Social Media+ Society*, 9(1), Article 20563051221147325

⁸² See the previously cited Verma, K., Davis, B., & Milosevic, T. (2022). Examining the Effectiveness of Artificial Intelligence-Based Cyberbullying Moderation on Online Platforms: Transparency Implications. *AoIR Selected Papers of Internet Research*.

⁸³ Suzor, N. P., West, S. M., Quodling, A., & York, J. (2019). What do we mean when we talk about transparency? Toward meaningful transparency in commercial content moderation. *International Journal of Communication*, 13, 18; Bates, J., Kennedy, H., Medina Perea, I., Oman, S., & Pinney, L. (2023). Socially meaningful transparency in data-based systems: reflections and proposals from practice. *Journal of Documentation*. Retrieved from: <https://www.emerald.com/insight/content/doi/10.1108/JD-01-2023-0006/full/html>

⁸⁴ Such as when VLOPs engage consultancy companies to conduct audits for them: See e.g. <https://about.fb.com/news/2020/08/independent-audit-of-enforcement-report-metrics/>

Appendix - We here incorporate research-based evidence for questions raised in Appendix 2. While we were not able to consult children specifically about the questions the Commission raised in Appendix 2, we provide answers therein based on our previous research with children and young people.

Q1. What do you like about being able to watch or share videos on websites or apps?

Our studies confirm that the most common social networks among post primary students in Ireland are those with video sharing features, as they are most frequently registered on YouTube, TikTok and Instagram, and only 1.9% reported not using any social network.⁸⁵ Other European research has shown that students appear to transfer their enjoyment for certain offline activities to the Web, as they want to interact with other people and express their own identities, reporting they feel less lonely and connect with people with similar interests on social media.⁸⁶ Social networking sites may have a positive impact on the well-being of users as they can increase perceived social support.⁸⁷

Q2. How safe do you feel when you are watching or sharing videos on websites or apps?

According to recent research from the EU Kids Online network, 28% of 9-16-year-old Internet-using children in Europe always feel safe online and 37% report that they often feel safe online.⁸⁸ Online risks do not necessarily turn into harm, and access to the Internet and usage of the mobile phone may bring adolescents benefits that make the risk-taking worthwhile (see our reply to Question 2 and Q9 for more detail).

Q3. Are you concerned about any videos that you see on websites or on apps? If you are, what types of videos concern you the most?

In our prior research specifically about witnessing cyberbullying, several participants expressed concern about targets and acknowledged the negative consequences online victimisation and cyberbullying can have, while others believe bullying is not so serious when it happens online.⁸⁹ There is even a minority advising to simply ignore the cyberbullying without taking further measures as the online environment is toxic by nature, but most would advise other people to minimise the risks and potential harm even if full prevention feels impossible, therefore being somewhat concerned but still choosing to engage online, probably

⁸⁵ Feijóo, S., Sargioti, A., Sciacca, B., & McGarrigle, J. (2023). Bystander Behaviour Online Among Young People in Ireland. DCU Anti-Bullying Centre, Dublin City University.

⁸⁶ Feijóo, S. (2022). *Problematic Internet Use and online risk behaviors. An analysis from the gender perspective*. Universidade de Santiago de Compostela.

⁸⁷ Verduyn, P., Ybarra, O., Résibois, M., Jonides, J., & Kross, E. (2017). Do social network sites enhance or undermine subjective well-being? A critical review. *Social Issues and Policy Review*, 11(1), 274–302.

⁸⁸ Smahel, D., Machackova, H., Mascheroni, G., Dedkova, L., Staksrud, E., Ólafsson, K., Livingstone, S., & Hasebrink, U. (2020). *EU Kids Online 2020: Survey results from 19 countries*.

⁸⁹ Feijóo, S., Sargioti, A., Sciacca, B., & McGarrigle, J. (2023). Bystander Behaviour Online Among Young People in Ireland. DCU Anti-Bullying Centre, Dublin City University.

because the benefits they obtain from risk engagement online outweigh the costs.

Q4. Do you feel that you have enough control over the type of videos that you see on websites or apps?

Young people in Ireland are mostly aware of the privacy settings and block buttons yet is unclear if they are fully aware of the possibilities these features give to filter the content they are exposed to,⁹⁰ and the social comparisons through SNS can lead adolescents to believe that their lives are a failure because they do not achieve the perceived standard.⁹¹ Therefore, there may be a mismatch between perceived control and the impact that content on websites and apps is having on young people.

Q5. Do you think that companies who run websites or apps that allow videos to be watched or shared should do anything to make things safer for you or your friends or family?

Several post-primary students in Ireland called for the facilitation of in-app reporting and to be provided a prompt response to the reported situation.⁹² Automatic word blocking was also raised as a possibility, as well as including a scoring system for users so their online reputation would warn others about their potential tendency to create content that is offensive or insulting to other users. It should also be noted that some young people wanted to be able to search for potential solutions offline, such as receiving training and involving schools in handling cyberbullying. The training requested includes information on how to act, emotional intelligence to understand the consequences of bullying, and developing self-confidence to be able to act when witnessing someone being mistreated online or offensive content posted. Teachers stand out as an important figure in this context, with students reporting the need to have a teacher designed to handle online incidents in their schools. In this sense, students seem to not fully distinguishing between offline and online incidents,⁹³ which conflicts with teachers not feeling that latter to be problems they should address as they perceive cyberbullying and other online issues to be incidents outside of school.⁹⁴ This would imply the need to first raise teacher awareness of online incidents as their responsibility and training to handle them.

⁹⁰ Ibid. footnote 79.

⁹¹ El Asam, A., Samara, M., & Terry, P. (2019). Problematic Internet use and mental Health among British children and adolescents. *Addictive Behaviors*, 90, 428–436.

⁹² Ibid. footnote 79.

⁹³ Pichel, R., Foody, M., O'Higgins Norman, J., Feijóo, S., Varela, J., & Rial, A. (2021). Bullying, Cyberbullying and the Overlap: What Does Age Have to Do with It? *Sustainability*, 13(15), Article 8527.

⁹⁴ Green, V. A., Johnston, M., Mattioni, L., Prior, T., Harcourt, S., & Lynch, T. (2017). Who is responsible for addressing cyberbullying? Perspectives from teachers and senior managers. *International Journal of School & Educational Psychology*, 5(2), 100–114.

Q6. How old do you think a child should be before they should be allowed to watch or share videos on websites or in apps? Should there be different rules for children who are different ages?

Actual content rating, like the Pan European Game Information (PEGI) and the Entertainment Software Rating Board (ESRB) have different thresholds of recommendations depending on age, but decision-making rests ultimately with the parents. All variables must be taken into account, including the level of maturity and other personal characteristics of each child, something that nor electronic system nor law can do.⁹⁵

Q7. Have you ever reported your concerns to your parent/s or guardian/s or to a company in charge of websites or apps about a video that you have seen? How did that go?

In our recent report *Bystander Behaviour Online Among Young People in Ireland*,⁹⁶ it was found that 56% of the participants who reported that they witnessed mistreatment online stated that they told someone about this experience, the preferred persons to talk about witnessing cyberbullying were parents/guardians and friends. We know from prior literature that children may be reluctant to report incidents on themselves as parents sometimes decide on turning off the computer or mobile as a solution, and this will cause their children to miss out on all the benefits the internet has to offer as well as being cut off from their friends.⁹⁷

Q8. Is there anything else you would like to comment on?

Internet and smartphone usage and the potential harms associated with their misuse need to be understood in a broader context. Factors such as parental supervision, depressive symptoms, overall psychological well-being, or the COVID-19 lockdown, have shown to play a role in whether and to what extent a child will develop a problematic use or engage in risky behaviour online.⁹⁸ Despite potential threats and harms online, that young people mostly report feeling positive emotions while online and have positive perceptions and expectations of their Network usage, implies that risk-taking is worth it for them and the online setting is fulfilling needs that would require to be addressed otherwise for prevention to be effective.⁹⁹ A holistic approach should also be adopted, rather than focusing on specific risky behaviours.¹⁰⁰ Increasing online safety by improving the digital skills of young people is necessary, but not sufficient, given the

⁹⁵ Mora-Salgueiro, J., Feijóo, S., Braña, T., Varela, J., & Rial, A. (2022). Gaming habits and symptoms of video game addiction. *Behavioral Psychology*, 30(3), 627–639.

⁹⁶ <https://antibullyingcentre.ie/wp-content/uploads/2023/05/Bystander-Behaviour-Online-Report.pdf>

⁹⁷ Sabella, R. A., Patchin, J. W., & Hinduja, S. (2013). Cyberbullying myths and realities. *Computers in Human Behavior*, 29(6), 2703–2711.

⁹⁸ Feijóo, S. (2022). Problematic Internet Use and online risk behaviors. An analysis from the gender perspective. Universidade de Santiago de Compostela.

⁹⁹ Milosevic, T., Kuldass, S., Sargioti, A., Laffan, D. A., & O'Higgins Norman, J. (2022). Children's Internet use, self-reported life satisfaction, and parental mediation in Europe: An analysis of the EU kids online dataset. *Frontiers in Psychology*, 12, Article 698176; Milosevic, T., Bhroin, N. N., Ólafsson, K., Staksrud, E., & Wachs, S. (2022). Time spent online and children's self-reported life satisfaction in Norway: The socio-ecological perspective. *New media & society*, Article 14614448221082651.

¹⁰⁰ Feijóo, S., Foody, M., O'Higgins Norman, J., Pichel, R., & Rial, A. (2021). Cyberbullies, the Cyberbullied, and Problematic Internet Use: Some Reasonable Similarities. *Psicothema*, 33(2), 198–205.

anticipated positive outcome derived from the online experience. Emotional and social skills, such as empathy, assertiveness and self-regulation, also need to be improved.¹⁰¹ Furthermore, research shows that at post-primary level risk behaviours are already established, showcasing a need for earlier prevention.¹⁰² Further recommendations include following a mainstreaming gender approach with a strong psychosocial component, taking into account the differences in which adolescents relate to each other online to the Net itself, and the divergence in terms of needs and emotions that are based on stereotypes and traditional gender roles. Consequently, although part of the prevention should be universal, it may be necessary to develop specific modules developed considering the needs of the different genders and moving away from the current androcentric approach.¹⁰³ Finally, a community-based approach is also recommended, with efforts at different social levels.¹⁰⁴ Families should be the main source of education in the safe and healthy use of technology, so parents and/or legal guardians may need to increase their own digital skills in order to better support and supervise their children in the online experience. At the same time, educators and schools need to play a more active role. Beyond the scarcity of time and resources and the urgency to meet the academic objectives set in the curricula, school is an excellent scenario of opportunity to include transversal content to improve the digital and social skills of young people. Besides, schools are the natural context for interaction with peers and the place where adolescents spend most of their time every day. This is related to the whole-education approach,¹⁰⁵ whereas all school staff, policy makers and other educational stakeholders and not only teachers are involved in prevention, but it also calls for other organisms to be involved. Other institutions and the public administration itself can provide resources and programmes on a continuous basis over time, encouraging activities outside the Internet by offering alternative leisure, sport and physical activities and promoting a regulatory framework that protects underage people in the online environment, such as the Online Safety Code.

¹⁰¹ Rial, A., Golpe, S., Isorna, M., Braña, T., & Gómez, P. (2018). Minors and problematic Internet use: Evidence for better prevention. *Computers in Human Behavior*, 87, 140–145.

¹⁰² Ibid. footnote 97

¹⁰³ Baxter, A., Salmon, C., Dufresne, K., Carasco-Lee, A., & Matheson, F. I. (2016). Gender differences in felt stigma and barriers to help-seeking for problem gambling. *Addictive Behaviors Reports*, 3, 1–8.

¹⁰⁴ UNICEF (2020). *COVID-19 and its Implications for Protecting Children Online*. UNICEF.

¹⁰⁵ O'Higgins Norman, J., Berger, C., Yoneyama, S., & Cross, D. (2022) School bullying: moving beyond a single school response to a whole education approach. *Pastoral Care in Education*, 40(3), 328–341.

Online Safety Codes for Video-Sharing Platform Services (VSPS): Submission to Coimisiún na Meán September 2023

Introduction: Safe Ireland

Safe Ireland is the national development and co-ordination body working to eradicate Domestic Violence (DV). We have five distinct functions: investigating the causes and effects of violence and coercion based on sex, gender and sexuality; delivering frontline refuge, support and outreach services; supporting the development, delivery and coordination of frontline Domestic Violence member services; developing best practice guidelines for skilled community-led domestic violence response; and influencing civil society and national strategic policy. This is achieved through our collaborations with our network of affiliated independent frontline DV services; local communities; professionals; public bodies; academic institutions; philanthropists; and corporate partners.

There are thirty-eight DV services across Ireland affiliated as members to Safe Ireland. Each deliver various combinations of services including national and local crisis helplines, emergency accommodation, housing and practical supports, one-to-one emotional and therapeutic support, information and advocacy, Garda / Court accompaniment, and Welfare advice. Twenty of these services operate staffed DV Refuges.

Our core strategic focus is to change culture, transform responses to sex, gender, and sexuality-based coercion and violence in communities across Ireland, and to progress towards creating a free and Safe Ireland for women, for young people, and for children.

Introduction: This Submission

Safe Ireland welcomes very much this opportunity to make submissions on the content and structure of the first draft Online Safety Code on Video-Sharing Platform Services (VSPS). Online safety is a particular concern because of the ever-increasing incidence of online forms of abuse in close relationships being reported to us by the women and children supported by our member services. In our view, online safety codes for video-sharing platform services, as for other internet services, should provide a clear and where necessary, detailed procedure through which online abuse may be excluded or where detected, taken down or otherwise deactivated as simply and quickly as possible. Where harmful online content remains available, it has enormous potential to cause further serious harm to victims of domestic abuse who may be already vulnerable in other ways. As much as possible should be done to minimise the risk of such harm, and the creation of a clear and workable Online Safety Code is a vital tool in this regard.

This Submission will follow the same question order and numbering as in the consultation document published by Coimisiun na Mean¹, but not all questions will be answered, only those most relevant to online safety as it affects women and children living with the effects of domestic violence and abuse. Where the questions are grouped together on a single theme, Safe Ireland's response may also address all the questions together. The questions set out below are all in a **different font** so that they are immediately obvious.

Our responses are informed by Article 28 (b) of the revised Audio-Visual Media Services Directive (2010/13/EU) as amended by EU Directive 2018/1808 as well as by the relevant provisions of the Online Safety and Media Regulation Act 2022² ("the Directive")³. With regard to types of harm, this Submission does not address the grave harms caused by material promoting self-harm, suicide, or behaviour characterising an eating or feeding disorder, nor does it concern itself with commercial communications.

3. Online Harms

Question 1: What do you think our main priorities and objectives should be in the first binding Online Safety Code for VSPS? What are the main online harms you would like to see it address and why?

Question 2: What types of online harms do you think should attract the most stringent risk mitigation measures by VSPS? How could we evaluate the impact of different types of harms e.g., severity, speed at which harm may be caused? Is there a way of classifying harmful content that you consider it would be useful for us to use?

Safe Ireland Responses to Questions 1 and 2:

- The Commission's main priorities and objectives should be the prevention of harm to the extent that this is possible through risk assessment and safety by design measures and through education, in the first place and wherever harmful online content has been included on any VSPS, the effective creation, monitoring and enforcement of practicable Online Safety Codes whose main objective is the mitigation of harm through accurate identification and swift take-down processes which are simple to use;
- The main online harms which in Safe Ireland's view to be addressed by the Commission and by the VSPS **should include** all 4 types harms listed by Article 28 (b) of the revised Directive and also those additional ones included in Part 11 of the Online Safety and Media Regulation Act 2022 (OSMR), namely online content which bullies or humiliates

¹ "Call for Inputs: Developing Ireland's First Binding Online Safety Code for Video-Sharing Platform Services" (July 2023), accessible online via this web-link: [CallForInputs_vFinal.pdf \(cnam.ie\)](#)

² Online Safety and Media Regulation Act 2022: full text in original form is accessible online at: [Online Safety and Media Regulation Act 2022 \(irishstatutebook.ie\)](#)

³ Accessible via this web-link: [CL2010L0013EN0010010.0001.3bi_cp 1..1 \(europa.eu\)](#) (consolidated text)

another person which gives rise to a risk to someone's life or a risk of significant harm to a person's physical or mental health, where the harm is reasonably foreseeable. In our view, the Commission should also include in its Online Safety Code definitions of harm – any harm which may be caused by the availability to minors of what is defined as “age-inappropriate online content” in Section 139D of the Broadcasting Act 2009 as inserted in Part 11 OSMR.

- We also think that all content consisting of material which it is already a criminal offence to disseminate, as defined in Schedule 3 OSMR, should be included in the definition of harmful online content, as its reach is wider than that of the corresponding subparagraph of Article 28 (b) of the Directive.
- With regard to age-inappropriate online content, our view is that it causes particular harm to the young but is not in fact appropriate to any age, as a very large proportion of it is violent in nature. It seems to us that easy access to material which models and glorifies violence in relationships is a very significant factor in the prevalence of domestic violence and abuse, and for this reason, this kind of material if it is to be made available at all, should not be available to children or young people, at least not to those under 18.
- In short, because domestic abuse is now being perpetrated in myriad online forms, because its effects can be very serious and long-lasting, and because the range of harms caused by online harmful content in the context of a close relationship is wide, it is appropriate to define harmful online content as widely as possible.
- Safe Ireland's view is that **all** of these types of harmful online content should attract stringent risk mitigation measures from VSPS and the Commission (through the Codes).
- With regard to the impacts of harmful online content, we are particularly mindful of the duty imposed by the Directive in Article 28 (b) on VSPS to protect minors from material “which may impair their physical, mental or moral development”. In our view, what is now defined in OSMR as “age-inappropriate content” may very easily impair the physical, mental or moral development of young people and therefore, should be as stringently controlled as other forms of material which are explicitly identified as “harmful online content” in our own OSMR.

4. Overall Approach to the Code

Question 4: What approach do you think we should take to the level of detail in the Code? What role could non-binding guidance play in supplementing the Code?

Safe Ireland Responses:

- Safe Ireland's view is that it may be most useful to go with Option 3 as set out at paragraph 4.1 of the Call for Inputs document – a mixed approach, that is, a brief statement of a high-level obligation, immediately followed by a statement of the

discrete concrete steps to be taken by the VSPS to comply with that obligation. These steps will need to be more detailed in some areas than others.

- With regard to non-binding online safety guidance materials, Safe Ireland’s view is that in the context of domestic violence and abuse, detailed guidance is needed on its nature, dynamics and impacts. A dominant theme in such guidance should be the importance of **context**: a single piece of online material which could be relatively harmless in another context might be very damaging in the context of a pattern of domestic abuse. Safe Ireland would be happy to be consulted further on the contents of such guidance at a later time, but for now the following outline of the main topics which we think ought to be included in such guidance is set out below:
 - The many-faceted nature of online abuse including within intimate and other close relationships;
 - The importance of context in deciding whether material falls into the “bullying and humiliating” category or not;
 - The multi-faceted and serious impacts of online interpersonal including sexual, forms of harassment and abuse;
 - The factors which make people vulnerable to various forms of harassment and abuse in close relationships;
 - The factors which may make it difficult for a victim to address such abuse;
 - The victims’ experience of online sexual and non-sexual harassment and abuse – their fears of continuing abuse, their loss of privacy, their damaged relationships - and their need for speed, simplicity, default takedowns, good communication, easy to understand information and advice, and for transparency in all processes;
 - Victims’ need to be treated with compassion and respect in circumstances in which they must relate intimate, embarrassing and frightening experiences to strangers.

Question 5: What do you think would be the most effective structure for the Code? What are the most important factors we should consider when we decide how to structure the Code?

Safe Ireland Responses:

- Safe Ireland’s view is that it would be most useful to combine the first two approaches listed in the Call for Inputs document at paragraph 4.2, that is, to have separate sections in the Code for each main category of content addressed, and then, to structure the rest of the Code thematically, more or less following the structure set out in Article 28(b).3 of the Directive, but perhaps grouping the ones related to Prevention together instead of in the order in the document, thus: Content Policies/Terms and Conditions, Online Safety Features, Service Design Measures, Risk Assessments, Content Moderation and Complaints and Compliance Measures. In our view, users and moderators employed by VSPS as well as Commission staff will all need to be able to access material on each main category of content readily, so that they can themselves identify it accurately and therefore, know what they are dealing with before they begin to address how to deal with it (by removing it, making a complaint, etc).
- We also suggest that it might be helpful to summarise processes in one or more flow-charts and to use colourful diagrams to summarise the differences between the various forms of harmful online content.
- Safe Ireland’s view is that it is important whatever structure is chosen that it is clear and logical and easy to navigate – using colour coding and coloured tabs to distinguish different processes/forms of harmful online content, for instance.

Question 6: How should we design the Code to minimise the potential for conflict and maximise the potential for synergies in how platforms comply with it and the DSA⁴?

Safe Ireland Responses:

- In Safe Ireland’s view, the approach taken by the DSA should be implemented in the Code as far as legally possible without creating a conflict with OSMR, not least because it has such a strong focus on the “privacy, safety and security of minors”. As far as minors are concerned, we approve strongly of the suggestion that there should be more specific requirements in areas such as age verification/assurance, content rating, and parental controls.
- With regard to the DSA obligation on VSPS to allow interested parties to notify them of illegal content and require them to take it down on receipt of a notice, we also approve of the suggestion that this should be extended to cover other types of harmful online content, in particular those we have identified above.
- Finally, we think that the obligation on VSPS in the DSA to process notices in a timely and diligent manner is a vital one which should most certainly be

⁴ Digital Services Act (EU) – it will come fully into effect in Ireland in February 2024. The full text is accessible online via this web-link: [Publications Office \(europa.eu\)](https://publications-office.europa.eu)

incorporated into the Code. It should be accompanied by obligations on VSPS to collect and on request, to provide data to the Commission on processing times and the accuracy of decisions taken on content.

Question 7: To what extent, if at all, should the Code require VSPS providers to take measures to address content connected to video content?

Safe Ireland Response:

- Safe Ireland’s view on this point is that the Code should require VSPS providers to take in essence the same measures to address content connected to video content (captions, blurbs, comments, voice-overs, sub-titles, etc) as they must with regard to the video content itself. Sometimes the entire harm lies in the caption or sub-titles accompanying the video itself. We see no reason in principle to distinguish between the video content and content connected to it, but content connected to it should be defined clearly and unambiguously.

5. Measures to be taken by Video-Sharing Platforms

Question 9: How should we ask VSPS providers to introduce and design a flagging mechanism in the Code? How can we ensure that VSPS providers introduce the mechanism in a user-friendly and transparent way? How should we ask VSP Providers to report the decisions they’ve made on content after it has been flagged? To what extent should we align the Code with similar provisions on flagging in the DSA?

Safe Ireland Responses:

- Safe Ireland’s view is that it is appropriate to follow the approach taken by the DSA at its Article 16 when designing the mechanism to be followed by users when reporting or flagging harmful content and imposing related obligations on VSPS. It makes no sense for users who are disturbed by a particular piece of content to have to determine first whether they should flag content under the DSA, or under the Code. All notifications should be processed in a timely, diligent, non-arbitrary and objective manner, as specified in the DSA.
- With regard to ensuring that the flagging mechanisms provided by VSPS to users are user-friendly and transparent, Safe Ireland’s view is that they should be always visible on screen, be written in simple language, contain the minimum of discrete steps to be taken and allow some means through which users with particular communications difficulties nevertheless can flag their concerns without difficulty.

- Safe Ireland’s view is that real-time data on numbers of notifications received, being worked on, decided upon and their classification (which type of harm and whether user is an adult, child or person with additional communication needs) should be collected and made available to the Commission continuously.

Question 10: What requirements should the Code include about age verification and age assurance? What sort of content should be shown by default to users who are logged out or in private browsing mode and whose age cannot be verified or assured? What evidence is there about the effectiveness of age estimation techniques? What current practices do you regard as best practice? Where accounts are not age verified should default privacy settings be used, should content default to universal content and should contact by others be more limited?

Safe Ireland Responses:

- In relation to “age-inappropriate content” as defined by Section 139D in Part 11 OSMR, we think VSPS should be required to take a very stringent approach to age verification procedures. Safe Ireland’s view is that this kind of material is harmful enough to justify the strictest possible measures: we suggest that there should be no access to this material without production of a driver’s licence or passport to those under the age of 18. If it is technically possible to supplement this with real-time processing of biometric data, this should also be done so that a young person cannot circumvent these controls by borrowing the passport etc of an adult. (We have in mind a process like the real-time passport photograph verification process used at some airports).
- Safe Ireland does not have access to evidence on the effectiveness of age estimation techniques. Our view is that verification should depend on harder evidence than anything that is estimated.
- Where accounts are not age-verified, our view is that the content should default to the kind of content that is suitable for the youngest users to minimise the risks to children and young people as much as possible.
- Finally, we are wary of any distinction being made in the Code between content which is suitable for older children but not younger ones. Older children do not have always have the maturity to prevent their younger siblings and friends from watching content which may be appropriate only for the older age group.

Question 11: What requirements should the Code have in relation to content rating? What do you consider to be best practice? What experiences have you had using content rating systems on platforms and do you think they have been effective? What steps could we ask VSPS to take to ensure content is rated accurately by users?

Safe Ireland Responses:

- Safe Ireland does not have the technical expertise to give a detailed answer to this set of questions. We will only offer general comments to the effect that:
 - This is a very subjective way of determining whether content is suitable, and judgments made often be made entirely in good faith by well-intentioned users with no expert knowledge of how harm might be caused by some types of content. It seems to us that it is not very useful to rely to any great extent on content ratings generated in this way, as the information may not be very accurate;
 - How would content ratings of any single video be synthesized? If the measure taken is a median rather than a mean, it would be a more accurate reflection of the overall content rating supplied by a number of people, but the problem of subjectivity remains;
 - Safe Ireland is concerned also about content ratings made in bad faith by bad actors with no concern for the welfare of minor users. We do not see how these could be very easily prevented except through extensive and continuous monitoring, which itself is not likely to be workable.

Question 12: What requirements should the Code have in relation to parental control features? How can we ensure that VSPS providers introduce the mechanism in a user-friendly and transparent way? Can you point to any existing example of best practice in this area? Should parental controls be ‘turned-on’ by default for accounts of minors or where age is not verified?

Safe Ireland Responses:

- Safe Ireland favours the turning on of parental controls by default for accounts of minor and where age is not verified, as a way of introducing “safety by design”.
- Parental control features should always be visible on any video via an eye-catching, easy to recognize prompt which leads onto to simply worded and brief instructions on how to impose these controls.

Question 13: What requirements should the Code contain to ensure that VSPS provide for effective media literacy measures and tools?

Safe Ireland Response:

- The Code should oblige the VSPS to provide easily accessible online information on media literacy measures and tools relevant to its own services on its website, written in clear and simple language and with versions available to users with particular communications difficulties (including younger children).

Question 14: How should we ask VSPS providers to address online harms in their terms and conditions in the Code, including the harms addressed under Article 28b? How should key aspects of terms and conditions be brought to users' attention? What examples are there of best practice in relation to terms and conditions including content moderation policies and guidelines?

Safe Ireland Response:

- We do not disagree with the list of prohibitions included in the Call for Inputs at paragraph 5.2.1, but we think it should also include material which bullies or humiliates another person, and which gives rise to a significant risk of harm to a person's physical or mental health which harm is reasonably foreseeable, and
- It should also include a prohibition on material which is age-inappropriate unless there are robust age-verification procedures in place and the user agrees not to make any attempt to circumvent or undermine these;
- In all these cases, we think the default penalty for transgression of these prohibitions should be account suspension or termination. They are serious enough to warrant this in terms of the potential harm which may be caused.

Question 15: How should we ask VSPS providers to address content moderation in the Code? Are there any current practices which you consider to be best practice? How should we address automated content detection and moderation in the Code?

Safe Ireland Responses:

- Safe Ireland's view is that it is challenging for moderators to make accurate decisions within as short a timeframe as possible. This may be particularly difficult where bullying or humiliating content has been uploaded as part of a much wider campaign of abuse and violence in a close relationship – the other examples of the abuse may only be found offline or on another platform.
- We suggest that this challenge might be addressed by inserting in the Code a requirement that moderators should ask the person making the complaint about this form of material to supply information about other aspects of the abuse, including examples of online abuse on other platforms – and also about the impact that the abuse has had on them, in sympathetic language explaining that the more information the person can give them, the less time it will take to make the decision and the more accurate it is likely to be.
- Safe Ireland also stresses once more that when dealing with violence and abuse in a close relationship the extreme importance of context in deciding whether material should be taken down or not.

- Safe Ireland also takes the point that borderline cases pose a challenge to moderators. In these cases, it seems to us that the strategy which carries the least risk of harm is the most appropriate one. We think referring these cases for a second opinion is a good idea, but of course, it is one that takes time, and meanwhile, the content is still on the platform. We suggest that in these borderline cases where a decision is not easily made and may require a second opinion, access to the content should be suspended pending a final decision.
- On timescales, we think there should be defined periods during which a decision must be made, but this should be subject to exceptions which would apply whenever a decision is particularly difficult and/or a second opinion must be sought. These exceptions would carry much less risk of harm to anyone if it were the rule in the Code that access to the content at issue is suspended pending a final decision. The reasons for the extended period and its likely length should be communicated clearly to the user without delay.
- Safe Ireland's view is that high-risk content should carry the greatest priority as that is the content which has the most potential for harm. It also makes sense for requests from other regulators to take priority over requests from others, provided those requests do not refer to high-risk content.
- With regard to automated content detection, we take the Commission's point that though generally more accurate than user-flagged complaints, it can make mistakes leading to enormous distress (and harm) being caused to those affected by those mistakes. We think the suggestion that priority should be given regardless of the kind of harmful online content (i.e., whether illegal or not) to notifications from "trusted flaggers" is an excellent one.

Question 16: What requirements should the Code include about procedures for complaint-handling and resolution, including out-of-court redress or alternative-dispute resolution processes? To what extent should these requirements align with similar requirements in the DSA? What current practices could be regarded as best *Commission on their complaint handling systems and what should those reports contain? Should there be a maximum time-period for VSPS providers to handle user complaints and if so, what should that period be?

Safe Ireland Responses:

- Safe Ireland's view is that as suggested in the Call for Inputs document at paragraph 5.2.3, there should be an integrated complaint-handling system covering both DSA and Online Safety Code matters because this would be more convenient both for users and VSPS providers.
- Safe Ireland also thinks that the channels for making complaints about VSPS content should be easy to access (visible on the platform at all times) and to use

(in simple language and containing the minimum of steps to be taken). Not all users making complaints wish to pursue their complaint online and for those who do not (perhaps because the person abusing them has access to all their devices) there should be clear signposting to other (especially offline) avenues of communication – physical address, SMS number on which to text, phone number through which a voicemail message might be left, e.g.

- There should be clear timelines for the processing of complaints and the time taken should be as short as possible, bearing in mind the need for accuracy. It should also be clear what the person making the complaint should do if this timeline is breached, and the circumstances in which exceptionally it may be necessary to exceed that timeline should be stated clearly.
- Safe Ireland’s view is that the ideal situation would be that the content which is the subject of the complaint should be made inaccessible online pending a decision on whether it should be removed, to minimise the risk of any harm resulting from its continued presence.
- Where any timeline indicated must be breached, the reasons for this should be stated clearly to the person making the complaint using the mode of communication which that person has indicated s/he prefers, and the proposed new timeline should also be indicated.
- In any case where it is possible that additional information from the person making the complaint might very well shorten the time taken to process the complaint, this should be clearly explained.
- Channels of communication between VSPS officials handling complaints and those making them should always be as secure as possible. It is advisable in any case with a background of violence or abuse in a close relationship to follow any indication from the person making the complaint as to which is the most secure channel from their point of view.
- Safe Ireland considers that it would be helpful in at least some cases for the person making the complaint to be able to access a single out-of-court dispute resolution body if s/he is not satisfied with the response from a provider’s internal complaint handling process.

Question 18: What approach do you think the Code should take to risk assessments and safety by design? Are there any examples you can point us towards which you consider to be best practice?

Safe Ireland Response:

- In Safe Ireland’s view, safety should be built into new platforms, programs, and applications to the greatest extent possible from the start (“safety by design”) and this should be enshrined in the Code, and they should all be subject to

detailed risk assessment as to all relevant kinds of harm, before going “live”. We understand that the professional body representing internet service providers in Ireland, Hotline.ie, itself favours this approach.

Question 19: How do you think that cooperation with other regulators and bodies can help us to implement the Code for VSPS?

Safe Ireland Response:

- The European Regulators Group for Audiovisual Media Services (ERGA) has access to enormous amounts of data from across many European countries. It has also considerable experience in finding ways to improve cross-border communication on media regulation matters. The Commission itself will be a digital services coordinator working under the DSA with other such coordinators in the EU and within the European Commission, and also within Ireland it is part of the Digital Regulators Group. These are opportunities for mutual learning and therefore, for the refinement and improvement of the Code over time as both the VSPS and the Commission gain experience of working with the new Code. It seems to Safe Ireland that it is important for the Commission to maximise its opportunities in this regard.

Question 22: What compliance, monitoring and reporting arrangements should we include in the Code?

Safe Ireland Response:

- It is extremely important that the VSPS themselves create and share information about the risks posed by their services with the Commission so that the Commission itself is able to assess the effectiveness of the measures which each VSPS has put in place to first assess and then eliminate as far as possible, these risks.
- Safe Ireland’s view is that the Commission is absolutely right to be more concerned about possible risks of harm to more vulnerable users such as children and young people (and we would add, adult victims of domestic abuse) when it comes to compliance, monitoring and reporting arrangements. We think that where VSPS has significant numbers of these more vulnerable users, it would be appropriate to require reporting on compliance regularly and perhaps more often than once a year. However, we also take the point that any annual (or more frequent) compliance statement from a VSPS should be approved by its Board of Directors, to ensure that it gets adequate internal scrutiny.
- Safe Ireland also believes that ad hoc assessments of compliance by a VSPS with the Code should be carried out by the Commission.

In conclusion, Safe Ireland is very willing to assist the Commission further to the extent that it can on any matter raised in this submission.

Safe Ireland

SI/LSM/Final

Dated this 1st day of September 2023

Contact: Caroline Counihan BL, Legal Support Manager

Email:

[REDACTED]

Mobile:

[REDACTED]

BFLGI

Baby Feeding Law Group Ireland

**BABY FEEDING LAW GROUP IRELAND'S INPUT FOR
DEVELOPING IRELAND'S FIRST BINDING ONLINE SAFETY
CODE FOR VIDEO SHARING PLATFORM SERVICES**

BACKGROUND

Baby Feeding Law Group Ireland (BFLGI) welcomes the opportunity to make a submission to inform a future consultation by Coimisiún na Meán (the “Commission”) on a draft Online Safety Code for Video-Sharing Platform Services.

BFLGI is an alliance of organisations and individuals working together to advocate for policies which protect the rights to food and health of all infants, young children, mothers, parents, and families by addressing practices that commercialise infant and young child feeding, threaten breastfeeding, and undermine good health.

BFLGI is part of a network including BFLG UK, Code Monitoring Northern Ireland, and the International Baby Food Action Network as well as a member of the Coalition 2030 alliance, who are working towards upholding Ireland’s commitment to achieving the Sustainable Development Goals (SDGs) by 2030.

Our members include individuals from academic disciplines including medicine, nursing, dietetics, public health, and law. We have representatives from national organisations such as the Association of Lactation Consultants of Ireland, Cuidiú, Friends of Breastfeeding and La Leche League Ireland.

BFLGI advocates to implement and enforce existing laws that relate to infant and young child feeding and health as well as campaigning for stronger legislation aligned with the International Code of Marketing of Breastmilk Substitutes and subsequent resolutions (‘The Code’).

The Code is a global public health policy designed to protect the general public, mothers, parents, and health professionals from the marketing practices of the baby food industry that have been shown to negatively impact breastfeeding practices. The objective of the Code is to protect and promote safe feeding practices for infants and young children. As part of this, and to protect the infant and young child feeding environment, the Code sets international standards for the safe promotion of all milk formula aimed at infants and young children aged up to 36 months (known as breastmilk substitutes (BMS) or commercial milk formula (CMF)).

The Code prohibits the marketing of all CMF. Ireland, as a WHO member state, has an obligation under the Code and international human rights law to embody the Code into domestic law. To date, Ireland has implemented laws that prohibit the marketing of CMF for babies up to 6 months but has failed to fully align with the Code to regulate the marketing of CMF up to 36 months (despite being an original signatory in 1981). Consequently, Irish mothers/parents are exposed to an extensive

range of CMF marketing. A contemporary problem within this is the growing threat of digital CMF marketing which gives companies unparalleled access to pregnant women, new mothers and parents. This allows the cross promotion of brands and undermines public health efforts and investment to support and protect breastfeeding.

BFLGI, and the Code, advocates for the protection of safe feeding for all children regardless of how they are fed. The medical advice for families who cannot, or choose not to breastfeed, is to feed first infant formula, along with complementary solid food from 6 months, until the infant is one year old and then switch to full fat cow's milk. The Department of Health, HSE and WHO all deem follow-on milks, toddler milks and other CMFs for babies older than 6 months "unnecessary".

The 2023 Lancet Series of Breastfeeding states the marketing of CMFs "comprehensively undermines access to objective information and support related to feeding of infants and young children. Additionally, CMF marketing seeks to influence normative beliefs, values, and political and business approaches to establish environments that favour CMF uptake and sales. In so doing, CMF marketing contributes to reduced global breastfeeding practices."

Digital platforms substantially extend the influence of marketing while circumventing the International Code of Marketing of Breastmilk Substitutes.

SUMMARY OF RECOMMENDATIONS

1. Harmful Commercial Communications, particularly marketing of commercial milk formulas and high fat, sugar and salt foods that undermine public health and infringe on fundamental rights as enshrined in the Convention on the Rights of the Child (CRC) (Article 24 “the right of the child to the enjoyment of the highest attainable standard of health) should be addressed.
2. The Online Safety Code must be prescriptive and high-level.
3. The Online Safety Code must ensure that children are protected effectively from harmful marketing and that their rights are upheld. This includes addressing commercial communications for mixed audiences, in order to capture all the marketing that children and their caregivers are exposed to.
4. The Online Safety Codes should protect all children, not just those old enough to have digital access. The Best Interests of the Child principle (under the CRC) determines that when an issue is being decided there is an intrinsic obligation on the State to consider the child’s best interests and this obligation is directly applicable so can be invoked before a court. Infants do not have capacity to advocate for rights, so rights are ascribed to them. The Online Safety Codes should protect the best interests of infants and young children as per the CRC by ensuring that those making decisions about their food and health are not exposed to harmful marketing.
5. The Online Safety Codes must address the heightened risks and harms associated with the commercialisation of infant and young child feeding and the negative impact on public health because of the marketing practices of CMF and HFSS food and drinks manufacturers.
6. Due to the risk of conflicts of interest, self-regulatory bodies should not be involved in the regulation of commercial communications or in the implementation of the Online Safety Code for VSPs.
7. The Commission needs strong, proactive enforcement mechanisms, which would apply punitive measures for instances of noncompliance and discourage infringements.
8. Continuous monitoring and enforcement mechanisms should be established (including a complaints procedure available to those with a legitimate complaint).
9. Enforcement mechanisms should be proactive and reactive - actively detecting infringements through monitoring and screening as well as accepting notifications of infringements.
10. The Commission should have clear authority to enforce the restrictions.
11. The Commission should be able to assess the effectiveness of procedural measures against a set of statutory objectives that go beyond simplistic content-related benchmarks such as removal rates and response times.

12. Regulated entities should not simply be required to provide periodic reports on their compliance or otherwise with codes but should also be compelled to provide any type of granular information to the Commission that is necessary for it to fulfil its supervisory tasks.
13. The Commission should have the power to demand any type of granular information that is necessary for it to fulfil its supervisory role. Shifting scrutiny towards these processes would help address some of the causal factors that give rise to harmful content online.
14. Child rights impact assessment (CRIA) should be mandated.
15. A dedicated function within the Media Commission should relate to online harms as they relate to data protection. As recommended by the Data Protection Commission, online harms that relate to data protection should be dealt with by the Media Commission.
16. Provision should be made to enable independent public interest research, based on data from platforms.

RESPONSES

The Call for Inputs document set out a number of issues and questions, exploring a wide range of topics. Questions relevant to the work of BFLGI, as well as associated groups and individuals, are addressed below in order. Some responses cover multiple questions, given some of the related content and to avoid duplication of responses.

Q 1: What do you think our main priorities and objectives should be in the first binding Online Safety Code for VSPS? What are the main online harms you would like to see it address and why?

Recommendations

Harmful Commercial Communications, particularly the marketing of CMFs, undermine public health, infringe on fundamental rights as enshrined in the Convention on the Rights of the Child and therefore should be addressed.

The Online Safety Codes must address the harms associated with the commercialisation of infant and young child feeding and the negative impact on public health because of the marketing practices of CMF and HFSS food and drinks manufacturers.

Priorities and Objectives

Article 5.1 of the International Code of marketing of Breastmilk substitutes specifies that there should be no advertising or other form of promotion of CMF to the general public. Yet across digital platforms, many digital marketing strategies are deployed. For example, data mining to identify and target pregnant women and mothers, the use of digital technologies to place promotions in direct response to concerns expressed online by pregnant women and mothers, promotion by influencers and social media platforms, the use of CMF brands' digital platforms to provide parents with parenting information.

To protect infant and maternal health and ensure the best interests of the child these types of commercial communications need to be strictly regulated.

The main objectives should be:

1. To provide a binding basis for a high level of public health protection in relation to commercial communications

2. To protect the fundamental rights of children and in particular their right to the enjoyment of the highest attainable standard of health, right to food, and right to privacy.
3. To uphold the best interests of the child as a primary consideration.

'Online safety' and 'online harms' should be defined broadly to include concerns related to digital marketing and data protection and privacy. Harmful digital marketing should be identified as a safety risk for children by States and by business actors themselves.

The 2020 WHO- UNICEF-Lancet Commission on the future for the world's children noted that "commercial marketing of products that are harmful to children represents one of the most underappreciated risks to their health and wellbeing".

Breastfeeding protects infants from life threatening infections, supports healthy brain development in children, prevents chronic childhood and maternal illness and reduces health care costs. CMF feeding increases incidence of infection, obesity, diabetes, and sudden infant death syndrome in infants. Some infants are particularly vulnerable when not fed human milk such as premature infants (due to the risk of necrotizing enterocolitis) and infants born into poorer socio-economic households. For mothers not breastfeeding is associated with an increase in certain cancers, type 2 diabetes, and heart disease.

Any regulatory scheme should be explicitly rooted in the international human rights framework. Children's rights are enshrined in the UN Convention on the Rights of the Child (UNCRC), which was adopted unanimously by the United Nations General Assembly in 1989 and signed up to by Ireland in 1992. It has since become the most rapidly and widely ratified human rights treaty in history, and its operationalisation is supported by a series of Optional Protocols and General Comments.

Both WHO and UNICEF recognise that the best way to protect children from harmful commercial communications, and respect, protect and fulfil children's rights, is to adopt a mandatory, comprehensive approach to regulation. Steps to restrict these harms must integrate both a public health lens and a child rights lens.

Children's digital rights have been an explicit concern of the international children's rights community. Accordingly, potential infringements to such rights must sit at the heart of considerations on online harms. Leading academics and experts in the area of law; child development; childhood studies; psychology; food and nutrition; media studies; and child, consumer, and digital rights call for the recognition of the far-reaching harms caused by digital marketing, the personal data extraction on which it is predicated, and the need to protect children from these in submissions to the UN

Committee on the Rights of the Child General Comment in relation to child rights in the digital environment.

This is because digital media marketing is subjecting children to intense commercial practices of implicit influence, neuromarketing, attitudinal structuring, and behavioural modification, without independent evaluation to ensure they do no harm. As a result, “children are thus commercial digital test subjects for marketing practices affecting their development, health and privacy.”

Online Safety Codes should build on the global initiatives underway at WHO, UNICEF, and other international agencies grounded in the fundamental rights of children. Enabling children of all ages to achieve their full developmental potential is a human right and a critical foundation for sustainable development. Children’s rights, including their rights to health, adequate and nutritious food, privacy, and to be free from exploitation, are threatened by commercial communications and their associated harms.

Q 3: Do you have reports, academic studies or other relevant independent research that would support your views?

Claim: Human milk is critical to the health and wellbeing of infants and young children as well as mothers, parents and societies

Supporting evidence

Emma Altobelli, Paolo Matteo Angeletti, Alberto Verrotti, and Reimondo Petrocelli. The Impact of Human Milk on Necrotizing Enterocolitis: A Systematic Review and Meta-Analysis (2020) 12 *Nutrients* 1

Olivia Ballard and Ardythe L Morrow, ‘Human Milk Composition: Nutrients and Bioactive Factors’ (2013) 60:1 *Pediatric Clinics of North America* 49.

Cesar G Victora, Rajiv Bahl, Aluísio Barros, Giovanny V A França, Susan Horton, Julia Krasevec, Simon Murch, Mari Jeeva Sankar, Neff Walker, and Nigel C Rollins ‘Breastfeeding in the 21st century: epidemiology, mechanisms, and lifelong effect’ (2016) 387 *The Lancet* 475.

[https://www.thelancet.com/journals/lancet/article/PIIS0140-6736\(15\)01024-7/fulltext](https://www.thelancet.com/journals/lancet/article/PIIS0140-6736(15)01024-7/fulltext)

Lars Bode, Arun S Raman, Simon H Murch, Nigel C Rollins, Jeffrey L Gordon, ‘Understanding the mother-breastmilk-infant “triad”’ (2020) 367 *Science* 1063.

Michael S Kramer and Ritsuko Kakuma, 'Optimal Duration of Exclusive Breastfeeding' (2012) 8 *Cochrane Database of Systematic Reviews* doi: [10.1002/14651858.CD003517.pub2](https://doi.org/10.1002/14651858.CD003517.pub2);

Nigel C Rollins, Nita Bhandari, Hajeobhoy, Chessa K Lutter, Jose C Martines, Ellen G Piwoz, Linda M Richter, Cesar G Victora, 'Why invest, and what will it take to improve breastfeeding practices?' (2016) 387 *The Lancet* 491
[https://www.thelancet.com/journals/lancet/article/PIIS0140-6736\(15\)01044-2/fulltext](https://www.thelancet.com/journals/lancet/article/PIIS0140-6736(15)01044-2/fulltext)

UNICEF 'The State of the World's Children 2019. Children, Food and Nutrition: Growing well in a changing world 2019' UNICEF (2019) [unicef.org/media/63016/file/SOWC-2019.pdf](https://www.unicef.org/media/63016/file/SOWC-2019.pdf)

Emily R. Smith, Lisa Hurt, Ranadip Chowdhury, Bireshwar Sinha, Wafaie Fawzi, Karen M. Edmond, on behalf of the Neovita Study Group, 'Delayed breastfeeding initiation and infant survival: A systematic review and meta-analysis' (2017) 12.

Andrew F Paquette, Beatrice E Carbone, Seth Vogel and Thomas Biederer, 'The human milk component myo-inositol promotes neuronal connectivity' (2023) *The Proceedings of the National Academy of Sciences* 120.

Kathleen M Krol and Tobias Grossmann, 'Psychological effects of breastfeeding on children and mothers' (2018) 61:8 *Bundesgesundheitsblatt - Gesundheitsforschung - Gesundheitsschutz* 977.

Dylan D Walters, Linh TH Phan, and Roger Mathisen, 'The cost of not breastfeeding: global results from a new tool' (2019) 34 *Health Policy and Planning* 407.

Claim: CMF marketing (seeks to) influence(s) the decision-making capacity of mothers, parents, caregivers and/or health professionals which is against the provisions of the Code and international law

Supporting evidence

Clare Patton and Amandine Garde, 'Harmful Marketing, Public Health and Human Rights Protection: the ongoing failure of Commercial Milk Formula companies to uphold their responsibilities under both the Code and the United Nations Guiding Principles' (forthcoming) *Frontiers of Public Health*.

Nigel Rollins, Ellen Piwoz, Phillip Baker, Gillian Kingston, Kopano Matlwa Mabaso, David McCoy, Paula Augusto Ribeiro Neves, Rafael Peres-Escamilla, Linda Richter, Kathryn Russ, Gita Sen, Cecilia Tomori, Cesar G Victora, Paul Zambrano, Gerard Hastings, Marketing of commercial milk formula: a system to capture parents, communities, science, and policy' (2023) 401 *The Lancet* 486.

WHO and UNICEF, How the marketing of formula milk influences our decisions on infant feeding (Geneva) 2022 [_Multi-country study examining the impact of BMS marketing on infant feeding decisions and practices, UNICEF, WHO 2022.pdf](#)

Gerard Hastings, Kathryn Angus, Douglas Eadie and Kate Hunt, 'Selling second best: how infant formula marketing works' (2020) 16:77 *Globalization and Health* 1 <https://pubmed.ncbi.nlm.nih.gov/32859218/>

Philip Baker, Julie P Smith, Amandine Garde, Laurence M Grummer-Strawn, Benjamin Wood, Gita Sen, Gerard Hastings, Rafael Pérez-Escamilla, Chee Yoke Ling, Nigel Rollins, David McCoy (on behalf of the 2023 Lancet Breastfeeding Series Group) *The political economy of infant and young child feeding: confronting corporate power, overcoming structural barriers, and accelerating progress.* (2023) 401 *The Lancet* 503. [https://www.thelancet.com/journals/lancet/article/PIIS0140-6736\(22\)01933-X/fulltext](https://www.thelancet.com/journals/lancet/article/PIIS0140-6736(22)01933-X/fulltext)

Changing Markets, 'Milking it: How Milk Formula Companies are Putting Profits Before Science' *Changing Markets Foundation* (2017)

changingmarkets.org/wp-content/uploads/2017/10/Milking-it-Final-report-CM.pdf

Claim: breastfeeding should be or is protected, promoted and supported in international law

Supporting evidence

Clare Patton 'Where do Human Rights Begin? In Small Places (*in the home*): Introducing a Right to Breastfeed and a Continuum Model of Duties to Protect Mother-Infant Dyads from Commercial Milk Formula Marketing' (*forthcoming*)

Amandine Garde, Seamus Byrne, Nikhil Gokani, and Ben Murphy, 'A Child Rights-Based Approach to Food Marketing: A Guide for Policy Makers' (UNICEF) (2018). sites.unicef.org/csr/files/A_Child_Rights-Based_Approach_to_Food_Marketing_Report.pdf

Amandine Garde, Seamus Byrne, Nikhil Gokani, and Ben Murphy, 'For a Children's Rights Approach to Obesity Prevention: The Key Role of an Effective Implementation of the WHO Recommendations' (2017) 8 *European Journal of Risk Regulation* 327

Helen Stalford, 'The Broader Relevance of Features of Children's Rights Law: The 'Best Interests of the Child Principle' in Eva Brems, Ellen Desmet, and Wouter Vandenhoele (eds), *Children's Rights Law in the Global Human Rights Landscape* (1st edn, Routledge, 2017)

Lida Lhotska, Veronika Scherbaum, Anne C. Bellows *Maternal, Infant, and Young Child Feeding. Intertwined Subjectivities and Corporate Accountability*, in Gender, Nutrition, and the Human Right to Adequate Food. Toward an Inclusive Framework 192 (Anne C. Bellows, Flavio I.S Valente, Maria Daniela, Nunez Burbano de Lara eds., 2016)

Judith Galtry, 'Strengthening the human rights framework to protect breastfeeding: a focus on CEDAW' (2015)

Benjamin Mason Meier and Miriam Labbok, 'From the Bottle to the Grave: Realizing a Human Right to Breastfeed Through Global Health Policy' (2009-2010) 60:4 *Case Western Law Review* 1073, at 1077.

George Kent, 'Child feeding and human rights' (2006) 44 *International Breastfeeding Journal* 93

Relevant international human rights law

Convention on the Rights of the Child (adopted 20 November 1989, entered into force 2 September 1990) 1577 UNTS 3 (CRC), Article 24 (2)(e).

UN Committee on the Rights of the Children, General Comment No. 14 on the right of the child to have his or her best interests taken as a primary consideration (art. 3, para. 1) (29 May 2013) UN Doc CRC/C/GC/14.

Committee on the Rights of the Child, General Comment 15 (2013) on the right of the child to the enjoyment of the highest attainable standard of health (art. 24)

Committee on the Rights of the Child (CRC), General comment No. 16 (2013) on State obligations regarding the impact of the business sector on children's rights.

Committee on the Elimination of Discrimination against Women, General Recommendation No. 34 (2016) on the rights of rural women, para 39(g).

Q 4: What approach do you think we should take to the level of detail in the Code? What role could non-binding guidance play in supplementing the Code?

Self-regulation offers no motivation to ensure meaningful commercial communications inline with global standards such as the Code to transnational corporations. This is evidenced in the fact that the CMF industry has spent the 40 years since the adoption of the Code developing loopholes to allow it to circumvent the provisions of the Code. The Online Safety Codes represent a tangible opportunity to shield Irish consumers from the proven exploitative online marketing that undermines public health and therefore the Online Safety Code should be prescriptive, detailed and sanctionable.

In 2022 the WHO's comprehensive analysis of the scope and impact of digital marketing of commercial milk formula found:

- Digital marketing is becoming the dominant form of marketing in many countries. In some countries more than 80% of exposure to breast-milk substitutes advertisements occurs online.
- Digital marketing increases breast-milk substitutes sales and occurs across multiple online channels and social media platforms in every country.
- Breast-milk substitutes companies buy direct access to pregnant women and mothers in their most vulnerable moments from social media platforms and influencers.
- Digital marketing can evade scrutiny from enforcement agencies and new approaches to implementing regulation and enforcement are required.

The sophisticated digital marketing deployed by CMF brands needs careful, considered and exacting monitoring and regulation and cannot be left to industry.

Q 6: How should we design the Code to minimise the potential for conflict and maximise the potential for synergies in how platforms comply with it and the DSA?

The Online Safety Code must be prescriptive and high-level.

EU legislation such as the General Data Protection Regulation, the Digital Services Act, as well as the Audiovisual Media Services Directive (the transposition of which is the basis for the development of this Online Safety Code) contain specific provisions related to child protection but most of them are principle-based and not concrete enough to be effective in practice without lengthy and costly litigation. Evidence shows that some major companies which are present in many children's lives are not sufficiently protecting them from online harms.

Commercial milk formula companies use strategies that are not immediately recognisable as advertising, such as online baby-clubs, advisory services, social media influencers, and user-generated content. Online due date calculators represent the very first privacy breach of a person - before they have even been born. Weeks after conception, companies hold data on expected babies, sell this data, and use it to serve bespoke, deep messaging to expectant mothers and parents. There is no way of knowing how long these companies might track the digital footprint of these infants.

Therefore, cross collaboration between DPC and OSC will need to address harms when personal data is used.

The new Digital Services Act contains provisions on protection of minors which are a step in the right direction. Article 28 requires 'appropriate and proportionate measures to ensure a high level of privacy, safety, and security' and a prohibition of displaying ads based on profiling using data from minors. However, it remains to be seen how platforms will effectively do this in practice.

Looking at existing legislative provisions and lack of detail in their implementation, no decisive approach currently exists to protect minors from harms caused by commercial communications. Therefore, the Code must be prescriptive and high-level.

Reliance on the development of codes of conduct that are not legally enforceable or subject to sanctions for non-compliance will not be sufficient.

Q10: What requirements should the Code include about age verification and age assurance? What sort of content should be shown by default to users who are logged out or in private browsing mode and whose age cannot be verified or assured? What evidence is there about the effectiveness of age estimation techniques? What current practices do you regard as best practice? Where accounts are not age verified should default privacy settings be used, should content default to universal content and should contact by others be more limited?

Age verification or limitations do not apply to babies and infants but the choices their parents make are shaped and manipulated by the digital environment. The exploitative marketing of commercial milk formulas targets caregivers. Parents are seeing harmful communications that affect the lives of their children.

Q21: Do you have any views on how requirements for commercial content arranged by a VSPS provider itself should be reflected in the Code?

Recommendations

The Online Safety Code must ensure that children are protected effectively from harmful marketing and that their human rights under the CRC are upheld. This includes addressing commercial communications for mixed audiences, in order to capture all CMF marketing.

While “women are the primary targets of formula milk marketing and have been for decades... Approaches aim to engage women early in their pregnancies to create brand loyalty from then through their children’s infancy, the toddler years and beyond” and these advertising strategies directly undermine children’s health and development.

Online Safety Codes should protect all children, not just those old enough to have digital access. Babies and infants are the most vulnerable children, and their protection should be extended through the caregiver by shielding the caregiver from infant formula marketing messages. The CRC identifies implementation of the International Code of Marketing of Breast-milk Substitutes and strengthening the State’s regulatory framework for industries and enterprises to ensure that their activities do not have adverse impacts on children’s rights as crucial steps to upholding the Convention on the Rights of the Child.

Q16: What requirements should the Code include about procedures for complaint-handling and resolution, including out-of-court redress or alternative-dispute resolution processes? To what extent should these requirements align with similar requirements in the DSA? What current practices could be regarded as best practice? How frequently should VSPS providers be obliged to report to the Commission on their complaint handling systems and what should those reports contain? Should there be a maximum time-period for VSPS providers to handle user complaints and if so, what should that period be?

Recommendations

The Commission should be able to assess the effectiveness of procedural measures against a set of statutory objectives that go beyond simplistic content-related benchmarks such as removal rates and response times.

The Commission should have the power to demand any type of granular information that is necessary for it to fulfil its supervisory tasks. Shifting scrutiny towards these processes would help address some of the causal factors that give rise to harmful content online.

Strong, proactive enforcement mechanisms are needed, which would apply stronger punitive measures for instances of noncompliance.

Tools and Resources

Proactive monitoring and enforcement are recommended to detect violations. Existing AI based technologies such as the CATCH tool can scan the internet to detect infringements. The WHO's NETCODE toolkit is a protocol to help establish a national ongoing monitoring system to assist governments in establishing a sustainable system that will monitor, detect and report violations of national laws and Codes. This enables relevant enforcement actions to be taken, so that violators can be held accountable for behaviour and practices that undermine breastfeeding and place the health of infants and young children at risk. For more information see:

<https://apps.who.int/nutrition/netcode/toolkit/en/index.html>

UNICEF provides direct legal assistance and expertise to governments and States to implement regulations that protect infant feeding and align with the International Code of Marketing of Breastfeeding Substitutes.

Transparency

The importance of transparency on the part of the services and platforms being regulated, and of the regulatory rules that are imposed on them, is important. Platforms and on-demand providers must respond to requests for information from the Commission.

Sufficient data to ensure thorough evaluations and independent public research is vital.

Complaint Handling and Self-Regulation

The era of self-regulation needs to end. Currently complaints regarding violations of national legislation pertaining to infant feeding marketing fall between the Food Safety Authority of Ireland and the Advertising Standards Authority of Ireland and are ineffective. Investigations occur only when complaints are made, and marketing has been seen by the public and judgements are made months after the marketing campaigns have ceased.

Problems with self-regulatory complaints mechanisms include:

- Complaint procedures do not provide a level playing field between citizens and industry: they are onerous and time-consuming processes for individual complainants.
- There is a lack of effective enforcement mechanisms such as fines to serve as a deterrent.
- Compliance and informal resolution processes are not open to public scrutiny.

The current enforcement mechanisms in place for non-broadcast commercial communications - of breaches being resolved by responding to individual complaints and promoting voluntary cooperation with the restriction – amounts to self-regulation, which has been shown to be ineffective and thus will not achieve the aim to minimise the harms associated with children’s exposure to commercial communications.

The failures of self-regulation, as well as the recommendations that the Media Commission will not cooperate with self-regulatory systems in the regulation of commercial communications and that non-statutory mechanisms are not considered as part of the regulatory framework, are covered in greater detail in the response to Question 19.

Q 18: What approach do you think the Code should take to risk assessments and safety by design? Are there any examples you can point us towards which you consider to be best practice?

Recommendation

Child rights impact assessment (CRIA) should be mandated. UNICEF and the WHO have recommended that in order to ensure that children’s best interests are adequately considered in food marketing restrictions, governments should consider carrying out an ex-ante child rights impact assessment (CRIA). CRIAs should help ensure that the best interests of children are taken into consideration during the policy and legislation development process and what the impact will be.

Indeed, the CRC states, “ensuring that the best interests of the child are a primary consideration in business related legislation and policy development and delivery at all levels of government demands continuous child-rights impact assessments” and, should therefore not be overlooked in the development of Online Safety Codes.

UNICEF’s Programme Guidance “Engaging with the Food and Beverage Industry” recommends no engagement with companies who violate the International Code of Marketing of Breastmilk Substitutes and no involvement of the food and beverage industry in public policy making.

Q 19: How do you think that cooperation with other regulators and bodies can help us to implement the Code for VSPs?

Recommendations

A dedicated function within the Media Commission should relate to online harms as they relate to data protection. As recommended by the Data Protection Commission, online harms that relate to data protection should be dealt with by the Media Commission.

Self-regulatory bodies should not be involved in the regulation of commercial communications or in the implementation of the Online Safety Code for VSPs

Priorities

- The objectives of addressing online harms on VSPs cannot be met in isolation without deep engagement with other regulators and consideration of interrelated issues, such as data protection, with the Data Protection Commissioner. The Online Safety Code should emphasise the extent to which online safety issues are interconnected with complex issues of data protection and privacy.
- Voluntary codes of practice should not be considered as a legitimate mechanism within the regulatory framework for online safety and should not be relied upon to stop harmful content online. Statutory mechanisms should be the sole structures by which Online Safety Codes are designed, implemented and enforced.

Data Protection

In the Submission by the Data Protection Commission to the Joint Committee on Media, Tourism, Arts, Culture, Sport and the Gaeltacht on the General Scheme of the Online Safety and Media Regulation Bill, the DPC referenced a regulatory lacuna and noted:

“In order to harness the full benefits of an Online Safety Commissioner as a constituent of the Media Commission and achieve meaningful outcomes for the public in this heretofore unregulated area, the DPC respectfully suggests that the Committee give due consideration to the following issues. Where a complaint or concern is raised about online content due to the harmful effects that content may have on the health/ safety/ wellbeing of one or more individuals, it should be dealt with through the regulatory framework envisaged by the OSMRB and via the enforcement powers of the Media

Commission (i.e. acting through an Online Safety Commissioner). It is possible that a single piece of content may be considered as falling within multiple categories of harmful online content, and the DPC believes that the possibility of such material also infringing multiple areas of law (including data protection) should be addressed within the OSMRB. Specifically, the DPC is strongly of the view that “material that violates [data protection or privacy law]” should absolutely not be excluded from the scope of harmful online content in Part 4.”

The DPC stressed that it was important that the Media Commission has the power to regulate all types of harmful online content, irrespective of whether they involve personal data. This is because there are clear limitations to the reach of data protection regulation, meaning it does not and cannot provide a comprehensive regime for tackling harmful content posted or shared in an online context.

While recognition is given in the Online Safety and Media Regulation Act legislation to harmonise some aspects of regulation of online safety that applies to data protection, the Online Safety Codes would benefit from a much more comprehensive understanding of the online harms related to data protection breaches through detailed explanation of how data protection online harms are to be addressed.

Indeed, while the Act provides that the Media Commission shall enter into memoranda of understanding with other relevant bodies, including the Data Protection Commission, there have been many criticisms levelled against the DPC on the capacity to fully and effectively execute its functions under the General Data Protection Regulation, with specific reference to its role as the Lead European Supervisory Authority in relation to large technology companies whose regional headquarters are located in Ireland. This has meant that the data rights of citizens of the European Union are being threatened.

While there is no intention for the Media Commission to supplant the role of the DPC in relation to data protection and privacy matters in any way, there must be a dedicated resource within the Commission that can be seconded to work on online harms as they relate to data protection. This is of particular importance given, as the Call for Inputs notes, “some of Europe’s largest VSPS providers are based in Ireland, and they provide large quantities of content to users in different languages and locations across the continent.”

Such overlap between the activities of the Media Commission and the DPC or potential synergies are already set to be addressed through a memorandum of understanding, which can be updated as needs be, but this additional resource can ensure that the burden of online harms pertaining to data protection is sufficiently addressed, especially if the DPC is overstretched.

Self and Co-Regulation

A 2013 systematic review found significant divergence between the reported impact of marketing regulation (including self-regulation by industry) provided in peer-reviewed journals, or industry-sponsored reports, showing the need for external monitoring.

Where those with commercial interests participate in the development and wording of self-regulatory codes, the resulting provisions are often so weak or unclear that they are meaningless. The "commitments" they contain, for instance, are often expressed as weak targets or goals, with thresholds so low that companies can reach them without much effort, and they routinely include imprecise wording which is open to interpretation.

A 2023 report on protecting children from the harmful impact of food marketing from the World Health Organization and the United Nations Children's Fund note that "the main stakeholders responsible for implementing effective policies to protect children from the harmful impact of food marketing should be trusted public authorities, as the bearers of a duty to protect children's rights and public health. Delegation of responsibility to other stakeholders (e.g., sector associations representing the advertising industry or broadcasters) is not recommended as it has been shown to create conflicts of interest at the heart of policy discussions in many countries."

Voluntary actions, such as industry-led pledges and other self-regulatory measures, have been demonstrated to be ineffective in protecting children from the impact of food marketing and commercial communications. They are not – and should not be viewed as – an appropriate mechanism to ensure that children are effectively protected from harmful marketing.

Furthermore, a child rights-based approach to the regulation of food marketing requires that competent public authorities do not engage in ineffective public-private partnerships amounting to the delegation of the mandate they have to protect child health and child rights to private business operators. and should therefore not be included as part of the regulatory package as part of the Online Safety Code.

Supporting evidence and research related to ineffectiveness of self-regulation includes:

Boyland, E.J. and Harris, J.L., (2017). Regulation of food marketing to children: are statutory or industry self-governed systems effective? *Public Health Nutrition*, 20(5), pp.761- 764.

Hawkes, C. (2008). Agro-food industry growth and obesity in China: what role for regulating food advertising and promotion and nutrition labelling? *Obesity Reviews*, 9, 151-161.

Kunkel, D. L., Castonguay, J. S., & Filer, C. R. (2015). Evaluating industry self-regulation of food marketing to children. *American Journal of Preventive Medicine*, 49(2), 181-187.

León-Flández, K., Rico-Gómez, A., Moya-Geromin, M. Á., Romero-Fernández, M., Bosqued-Estefania, M. J., Damian, J., ... & Royo-Bordonada, M. A. (2017). Evaluation of compliance with the Spanish Code of self-regulation of food and drinks advertising directed at children under the age of 12 years in Spain, 2012. *Public Health*, 150, 121-129.

Mackay, S. (2009). Food advertising and obesity in Australia: to what extent can self-regulation protect the interests of children. *Monash UL Rev.*, 35, 118.

Reeve, B. and Magnusson, R., (2018). Regulation of food advertising to children in six jurisdictions: a framework for analyzing and improving the performance of regulatory instruments. *Ariz. J. Int'l & Comp. L.*, 35, p.71

Sing, F., Mackay, S., Culpin, A., Hughes, S., & Swinburn, B. (2020). Food advertising to children in New Zealand: A critical review of the performance of a self-regulatory complaints system using a public health law framework. *Nutrients*, 12(5), 1278.

Thornley, L., Signal, L., & Thomson, G. (2010). Does industry regulation of food advertising protect child rights?. *Critical Public Health*, 20(1), 25-33.

World Cancer Research Fund International (2020). Building Momentum: lessons on implementing robust restrictions of food and non-alcoholic beverage marketing to children. Available at wcrf.org/buildingmomentum

Q 22: What compliance monitoring and reporting arrangements should we include in the Code?

Recommendations

On the issue of monitoring and enforcement, BFLGI endorses the processes and actions put forward by UNICEF and the WHO in terms of protecting children from harmful food marketing:

- The application of deterrent sanctions for non-compliance. Enforcement mechanisms should be both reactive and proactive, meaning that they should be open to both receiving notification of infringements, and detecting infringements through screenings and ongoing monitoring.
- Continuous monitoring and enforcement mechanisms should be established (including a complaints procedure available to those with a legitimate complaint)
- Clear authority to enforce the restrictions.
- Use of technology to proactively monitor the internet for infringements of online safety code.
- Use of existing protocol Netcode to ensure careful creation of online safety code.

Furthermore:

- Regulated entities should not just be required to “provide periodic reports on their compliance or otherwise with codes” but should also be forced to provide any type of granular information to the Commission that is necessary for it to fulfil its supervisory tasks.
- Provision should be made to enable independent public interest research, based on data from platforms.

Codes of conduct do not, by definition, include meaningful sanctions for those who do not comply with the code, or who are found to be in breach. The threat of “reputational damage” is incorrectly perceived as adequate deterrent for companies from breaching these codes.

Voluntary codes are particularly susceptible to breaches of all or some of their provisions when it is more commercially advantageous to do so. In the absence of sanctions for non-compliance, companies will continue to flaunt the code. This is especially true if there is not public awareness of the code or the complaints process. Appropriate sanctions must be set for non-compliance- It is not enough to rely on the censure of civil society and the media for failure to comply. Failure to comply with restrictions established through laws or regulations must lead to the application of effective sanctions.

Response to call for input: Ireland's First Binding Online Safety Code for Video-Sharing Platform Services

Introduction

The 5Rights Foundation welcomes the opportunity to comment on Ireland's First Binding Online Safety Code for Video-Sharing Platform Services (VSPS). The rights of the child, as established by the UN Convention on the Rights of the Child and elaborated as regards the digital environment in UNCRC General comment No. 25, must be a key element underpinning legislation in this space, both at EU and national level. In view of the special consideration of children's rights in the Digital Services Act (EU Regulation 2022/2065) and the Audiovisual Media Services Directive (EU Directive 2010/13/EU), as well as the Irish Data Protection Commission's Fundamentals for a Child-Oriented Approach to Data Processing, 5Rights believes that the Online Safety Code is an opportunity to foster the synergies between all the foregoing legislative and voluntary measures, thus improving their effectiveness and ultimately advancing the protection of children's rights online.

This document outlines 5Rights' key considerations and input on how the Online Safety Code can protect and promote children's rights in the digital environment. 5Rights develops policy, creates innovative frameworks, develops technical standards, publishes research, challenges received narratives and ensure that children's rights and needs are recognised and prioritised in the digital world. While 5Rights works exclusively on behalf of and with children and young people under 18, our solutions and strategies are relevant to many other communities. Our focus is on implementable change and our work is cited and used widely around the world. We work with governments, inter-governmental institutions, professional associations, academics, businesses, and children, so that digital products and services can impact positively on the experiences of young people.

3.1 What online harms should the Code address?

Question 1: What do you think our main priorities and objectives should be in the first binding Online Safety Code for VSPS? What are the main online harms you would like to see it address and why?

One in three internet users worldwide are children. The digital world is not optional for them. It is where they access education, health services and entertainment, build and maintain their relationships, and engage in civic and social activities. Children do not only use services explicitly targeted or designed for them, so they must be protected wherever they are in practice, not only where government, companies or parents and carers might wish them to be. A recent research on Children's Online User Ages revealed that 60% of children aged 8-12 use platforms whose minimum age is 13 with their own profile.¹ To protect children's rights, safety and to ensure their wellbeing, regulation must be geared towards all services that children access in reality. Therefore, the protection of children as well as the promotion of their rights, wherever and whenever they are online, must be a priority of the first Online Safety Code for VSPS, in compliance with UNCRC General comment No. 25 on children's rights in relation to the digital environment, which recognises the need to uphold all children's rights in the digital environment.

¹ Children's Online User Ages: Quantitative Research Study, OFCOM (2022), [link](#)

The Code should be comprehensive and cover all types of harms that children are exposed to. Under the 4Cs classification, children face four types of risks of harm online: content, contact, conduct and contract. Some of those harms are explicitly mentioned in article 28b of the revised Audiovisual Media Services Directives. All should be addressed under the Code.

Question 2: What types of online harms do you think should attract the most stringent risk mitigation measures by VSPS? How could we evaluate the impact of different types of harms e.g. severity, speed at which harm may be caused? Is there a way of classifying harmful content that you consider it would be useful for us to use?

Online harms that should attract the most stringent risk mitigation measures by VSPS are those impacting or putting at severe risk the most vulnerable categories of users, notably children. Children are significantly impacted by the omniscience of digital services in their life, as they generally spend several hours a day online. In particular, viewing videos is an almost universal activity for children.² Youtube is the most used platform amongst 3-17 years old.³ Research from Internet Matters shows a clear pattern between spending more time online and being more likely to experience an online harm.⁴ In addition, research shows that using VSPS children are particularly exposed to risks related not only to content but also to addiction to the services and related mental health issues – from disrupted sleep patterns to anxiety, depression and decrease in life-satisfaction.⁵

It is important to recognise that not all children are the same, and children's experiences online are shaped by multiple environmental and personal factors, including age, developmental capacity, gender, sexuality, and familial circumstances. It follows that there will be certain groups of children, at certain times, who will be more vulnerable to the effects of harmful content. Children will experience periods of increased sensitivity relative to their developmental stage, but not necessarily at the same age. In girls, for example, social media use between the ages of 11 and 13 years is associated with a decrease in life satisfaction one year later, whereas in boys this happens later between the ages of 14 and 15 years, suggesting that sensitivity to social media is linked to developmental changes that occur later in boys than girls.⁶

To evaluate the impact of harmful content, different factors should be taken into account:

1. Immediate and Cumulative

The impact of content can be immediate if the content is graphic, whereas content that is non-egregious in nature but can be harmful in large volumes may have a cumulative impact on children. A steady drip feed of body-image focused content accompanied by virality, and high engagement such as likes and comments can normalize unrealistic body types or sexualized images of young women and girls. According to Meta's own internal research, a third of teenage girls also believed that Instagram made them feel worse about their bodies.⁷ Molly Russell was 14 when she ended her life after viewing graphic self-harm, suicide and depression related content on social media. The coroner leading the inquest into Molly's death concluded that she had died from an act of self-

² Children and parents: media use and attitudes report 2023, OFCOM, [link](#)

³ Ibid; Teens, Social Media and Technology 2022, Pew Research Centre, [link](#)

⁴ Exploring the impact of online harms, Part 2, Internet Matters, [link](#)

⁵ 5Rights *Disrupted Childhood: the cost of persuasive design*, 2023

⁶ Windows of developmental sensitivity to social media, Orben, A., Przybylski, A.K., Blakemore, S.J. et al Nature Communications 13, 1649, [link](#)

⁷ Facebook Knows Instagram Is Toxic for Teen Girls, Company Documents Show, Wall Street Journal, [link](#)

harm while suffering from depression and “the negative effects of online content.”⁸ The coroner observed that while the content itself was harmful, it was made considerably worse by features such as comments, hashtags and likes, with some posts attracting over 10,000 likes.⁹ High numbers of likes and comments created a sense of legitimacy and normalised the extreme content, and as the coroner noted “glamorised and even glorified” self-harm.¹⁰ This illustrates how a large volume of pro-self-harm and pro-suicide content directed at or accessed by children can have a significant cumulative impact.

2. Acute and mild

Content may have an acute or mild impact on a child according to the child and their circumstances and characteristics. Factors such as gender must be taken into consideration when assessing the impact on content that is harmful to children. Many studies have found that the correlation between social media use and harm is stronger among girls. Meta’s own internal research, revealed by whistle-blower Frances Haugen, showed that among teenage girls experiencing suicidal thoughts, 6% in the US and 13% in the UK traced those thoughts back to Instagram.¹¹ As the divide between ‘online’ and ‘offline’ has become less distinct, the negative effects of children’s offline experiences carry over into their ‘online’ lives and vice versa. Content and interactions online that feel more personal, familiar or local can impact a young person’s ‘offline’ behaviour.

3. Direct and indirect

The way in which a child is interacting with the content can influence the content’s impact on them. If a child is directly viewing or experiencing the content or the content is about them the impacts will likely be more substantive. The digital environment in which the content is being experienced or generated in can also shape how direct or indirectly the impacts of harm are felt. A survey by the NSPCC and the Children’s Commissioner for England found that 44% of boys aged between 11 and 16 who regularly viewed pornographic content reported that it gave them ideas about the type of sex that they wanted to try.¹² Most young people also said girls expect sex to involve physical aggression, such as airway restriction.¹³ This corroborates findings from the UK school’s regulator as part of its review into sexual abuse in schools, which found that “children and young people... had learned more about sexuality from social media than from school or had got their education about relationships from their peers and social media.”¹⁴ Cumulative active engagement with hazards over time can self-reinforce behaviour to cause significant and severe harm such as engaging with and participating in pro-anorexia communities online. Cumulative passive exposure to hazards over time that can build up to cause more significant harm such as being immersed in body-focused content in social media feeds.¹⁵

Q3 Do you have reports, academic studies or other relevant independent research that would support your views? If you do, please share them with us with links to relevant reports, studies or research.

The following research reports all elaborate on the risks and/or harms that children can experience online, including when accessing VSPS.

- Disrupted Childhood: The cost of persuasive design (2023), available at <https://5rightsfoundation.com/uploads/Disrupted-Childhood-2023-v2.pdf>

⁸ REGULATION 28 REPORT TO PREVENT FUTURE DEATHS, Senior coroner Andrew Walker, [link](#)

⁹ Molly Russell Inquest proceedings, 2022

¹⁰ Molly Russell Inquest proceedings, 2022

¹¹ Facebook Knows Instagram Is Toxic for Teen Girls, Company Documents Show, Wall Street Journal, [link](#)

¹² ‘A lot of it is actually just abuse’- Young people and pornography, Children’s Commissioner for England, [link](#)

¹³ Ibid

¹⁴ Review of sexual abuse in schools and colleges, Ofsted, [link](#)

¹⁵ Research into risk factors that may lead children to harm online, OFCOM, [link](#)

- Pathways: How digital designs put children at risk (2021), available at <https://5rightsfoundation.com/uploads/Pathways-how-digital-design-puts-children-at-risk.pdf>
- But how do they know is a child? Age Assurance in the digital world (2021), available at https://5rightsfoundation.com/uploads/But_How_Do_They_Know_It_is_a_Child.pdf

4.1 How prescriptive or flexible should the Code be?

Question 4: What approach do you think we should take to the level of detail in the Code? What role could non-binding guidance play in supplementing the Code?

The Code should be future-proof, tech-neutral and outcomes-based, thus easing compliance by allowing all VSPS providers to constantly innovate and seek the best ways to comply with their obligations under the Code, including with regards to emerging technologies and types of harm. The Code should thus set out categories of harm for VSPS providers to address and oblige them to take appropriate measures to reduce the risks of harm and mitigate existing risks. These high-level obligations should be supplemented with more detail, notably via references to existing guidelines and the highest industry standards available at EU or national level.

4.3 How should the Code take account of the Digital Services Act (“DSA”)?

Question 6: How should we design the Code to minimise the potential for conflict and maximise the potential for synergies in how platforms comply with it and the DSA?

The Code could impose additional and more detailed requirements on VSPS providers, including with regards to the privacy, safety and security of minors that are specific for VSPS, for instance through more precise requirements related to age assurance, content rating, profiling of minors by recommender systems or parental controls. Such additional and/or more detailed requirements should however not conflict with or depart from guidance or specifications on the implementation of the DSA, but rather draw from the provisions of the DSA where possible, so as to ease compliance with all applicable rules. The Code should contain specifications on how video-sharing online platforms or very large online platforms can comply with the DSA with regards to video-sharing features, including via reference to voluntary framework on age-appropriate design of service (e.g. the Irish Data Protection Commission’s Fundamentals for a Child-Oriented Approach to Data Processing) and technical industry standards as appropriate.

5.1.3 Age Verification and Age Assurance Features

Q10 What requirements should the Code include about age verification and age assurance? What sort of content should be shown by default to users who are logged out or in private browsing mode and whose age cannot be verified or assured? What evidence is there about the effectiveness of age estimation techniques? What current practices do you regard as best practice? Where accounts are not age verified should default privacy settings be used, should content default to universal content and should contact by others be more limited?

Children have a right to access and participate in the digital world, and a right to access information which is not harmful to them and within the law¹⁶. Children must not be locked out of spaces they have a right to be in. Services must ensure they design features

¹⁶ UN Convention on the Rights of the Child, [link](#); UN General comment No. 25 (2021) on children’s rights in relation to the digital environment, [link](#)

and systems with children's safety in mind. If a service is found to pose risk to children, as a measure of last resort it may be necessary to restrict a child's access to a service or a part of a service. For instance, children should not be able to access commercial pornography services due to the high risk this material poses to children. A service which hosts age-appropriate content to children but does host some mature content should not lock children out of their service but must be able to prevent children from encountering only that content and activity that may be harmful to them.

Different age assurance¹⁷ systems provide a sliding scale of confidence in the user's age. Systems can be used on their own or in combination to ensure higher confidence in the result. Combining age estimation systems can lead to very high levels of confidence in the final result¹⁸. Some common reasons for using age assurance are likely to be:

- To prevent underage users purchasing age-restricted goods
- To prevent underage users accessing or procuring age-restricted services
- To prevent underage users viewing/accessing/consuming age-restricted content
- To provide age-appropriate experiences for different age groups

The level of assurance should be calibrated to the nature and level of risk presented by a product or service in relation to the age of the child. Crucially, age assurance must not be used to prevent children from participating in the digital world or to downgrade their experience. If a product or service is compliant with relevant data protection regulations, and is appropriate for children of any age, there may be no need for age assurance. In general, less risky services will require a lower level of assurance. Services presenting a high risk to children, where the likelihood of harm to children occurring is high, or the impact of the harm is not minimal, including services required to comply with legal age limits, will need the highest bar of assurance.

Common approaches to age assurance include several techniques and tools.¹⁹ Depending on the purpose, context and level of risk, services may implement a combination of approaches to age assurance.²⁰ A lack of minimum standards may lead to the exacerbation of known problems of excessive data collection,²¹ privacy infringements²², ineffective age checks²³ and could lead to heavy-handed age-gating that can block children out of spaces they have a right to be in. It is key that age assurance is used in a way that is proportionate to the level of risk on their services and abides by the below principles:

¹⁷ Age assurance is a mechanism with which services can ensure children are served age-appropriate experiences. Age assurance describes any system or feature which purports to estimate or verify the age or age range of a user. It is an umbrella term that captures a huge variety of approaches to ascertain age and encompasses both age estimation and age verification systems. Age verification systems determine a person's age with a high level of certainty by checking against trusted, verifiable data. Age estimation systems on the other hand estimate a person's age, using a combination of user provided data and algorithmic computation based on a large dataset. Outputs vary from a binary determination as to whether someone is or is not above or below a certain age, through to placing an individual in a specific age category, through to estimating an exact age. Cfr: *But how do they know it's a child?*, 5Rights Foundation, available at: https://5rightsfoundation.com/uploads/But_How_Do_They_Know_It_Is_a_Child.pdf

¹⁸ Ibid

¹⁹ Ibid: Self-declaration — requires a user to enter their birthdate or tick a box that asks if they meet the minimum age of use; Hard identifiers — requires users to provide verified sources of identification to prove their age, such as a passport; Biometrics — uses biometric information such as height, gait, voice, facial features, keystroke dynamics or finger and palm prints to identify a particular person or estimate their age; Profiling and inference models — uses data from user behaviour to infer the age of users; Capacity testing — estimates a user's age based on an assessment of their aptitude or capacity; Account holder confirmation — requires a parent to confirm the age of a child user; Device/operating system controls — offering controls on devices or through operating systems to deliver more age-appropriate experiences for children; Flagging — allows users to 'flag' other users they believe do not meet a service's age requirements

²⁰ Ibid

²¹ Man files complaint accusing YouTube of harvesting UK children's data, The Guardian, [link](#)

²² Largest FTC COPPA settlement requires Musical.ly to change its tune, Federal Trade Commission, [link](#)

²³ 60% of UK children aged 8-12 have a profile on at least one social media service, despite most social media having a minimum age requirement of 13, OFCOM, [link](#)

- Age assurance must be privacy-preserving;
- Age assurance should be proportionate to risk and purpose;
- Age assurance should be easy for children to use;
- Age assurance not unduly restrict access of children to services to which they should reasonably have access, for example, news, health and education services;
- Age assurance providers must offer a high level of security;
- Age assurance providers must offer routes to challenge and redress;
- Age assurance must be accessible and inclusive;
- Age assurance must be transparent and accountable;
- Age assurance must be rights-respecting.

While age assurance is a useful tool for serving children age-appropriate experiences and preventing them from encountering harmful content and activity, age assurance alone is not sufficient for making a service age-appropriate for children. Action should be taken to mitigate risk, taking into account the ages of users and the particular risks posed by the service. For example, to make a service age appropriate a service some providers might simply need to disable some of their more intrusive or risky design features. The best approach to age assurance will be dependent upon the nature of the service being provided, the users that access the service, the type of content and activity on the service and the way that policies and terms and conditions are set out. Any system of age assurance should not gather any more data than necessary about an individual to establish their age. Once that age or age range is established, the data used in the process should be stored or discarded transparently and securely. Whatever the technology, age assurance systems can be privacy preserving if operated in accordance with standards of data minimisation and purpose limitation. Children should not be routinely asked to disclose more information than is necessary to prove their age. Identity verification and age assurance should not be conflated. There is no need to confirm a user's identity when assuring their age.

5.1.5 Parental Controls

Q12 What requirements should the Code have in relation to parental control features? How can we ensure that VSPS providers introduce the mechanism in a user-friendly and transparent way? Can you point to any existing example of best practice in this area? Should parental controls be 'turned-on' by default for accounts of minors or where age is not verified?

Parental controls can be used to supplement a safety-and-privacy by design approach as they enable parents or carers to protect their child in a more individualised manner. It is essential that children maintain a certain degree of autonomy while using online services. As privacy is essential for children and teenagers in their development, parental controls should diminish following the child's evolving capacity. If parental controls are provided, it should be clear to the children that they are being monitored. For instance, if the online service allows a parent or carer to track their location or read their messages, an obvious sign must be given to the child. Furthermore, information on parental controls should be provided in an age-appropriate way detailing the data or activities that are being shared.

5.2 Terms and Conditions, Content Moderation and Complaints

Question 14: How should we ask VSPS providers to address online harms in their terms and conditions in the Code, including the harms addressed under Article 28b? How should key aspects of terms and conditions be brought to users' attention? What examples are there of best practice in relation to terms and conditions including content moderation policies and guidelines?

As recognised also by the Digital Services Act, Terms & Conditions should be concise, prominent and written in clear language suited to the age of children likely to access the service. To make published terms age-appropriate, organisations must consider the following:²⁴

Language: Published terms should avoid jargon and spell out key definitions and terms. The language used should be simple, straightforward and pitched at a level that the youngest likely user can understand, or presented in different versions to suit different age groups.

Length: Published terms should be concise and to the point. They should be short in length (word count), divided into clear sections or made available in bite-sized pieces.

Format: Key terms and definitions should be prominent, presented in bold text or graphics and icons if needed. Consulting with children on the most appropriate format and testing these with diverse groups of children enables their views to be heard and can guide formatting and design decisions.

Navigability: Published terms should be prominent and easy to find and key terms and definitions should also be searchable.

Timing: Published terms should be presented at multiple or significant times in the user journey. Ongoing, meaningful engagement at regular intervals and at crucial moments, including every instance where consent is sought, can support a child to comprehend any terms of agreement they are entering into.

Accessibility: Published terms should consider the diverse needs of young people. This includes providing terms in multiple languages and catering for children with accessibility needs. Providers should not assume children have an engaged adult on hand to help them understand terms. The following factors should be considered when making a product or service inclusive and accessible:²⁵

- The needs of children with disabilities
- The age or age range of the child
- The needs of children who may not have active or engaged parents or guardians
- The needs of vulnerable groups and children with protected characteristics
- The affordability of the product or service

Ensuring meaningful consent: Consent must be sought and obtained, not assumed, and must be given by a “clear affirmative act establishing a freely given, specific, informed and unambiguous indication.”²⁶ Obtaining meaningful consent means that a child understands and accepts to terms at all stages of their user journey and may choose to change their mind at a later point. ‘Tick box’ or ‘unread’ consent must not be used when

²⁴ Tick to Agree, 5Rights Foundation, [link](#)

²⁵ IEEE Standard for an Age Appropriate Digital Services Framework Based on the 5Rights Principles for Children, IEEE Consumer Technology Society, [link](#)

²⁶ What is valid consent, ICO, [link](#)

the end user is a child.²⁷ Children must be given the option to refuse individual terms without being precluded access to other parts of the service. Where parental consent is required, this should be meaningful and steps should be taken to verify that the parent or guardian is who they say they are. Providers should seek consent whenever amendments are made to the service, explaining the changes and their implications for the user, and it should be possible for users to withdraw consent, both after regular periods of time and at times of their own choosing.

Upholding published terms: Terms must be consistently enforced to create a culture of good governance and clarity for parents and young people about what constitutes a violation of service use agreements. Redress and reporting information must be prominent and easily accessible. It must also be clear what happens when a user makes a complaint. Expectations of response times must be clearly set out in the terms and upheld by the provider, and reports relating to young people's safety should take priority. Organisations should consult stage 12 of IEEE Standard 2089 for more details on making published terms age-appropriate.²⁸

Question 16: What requirements should the Code include about procedures for complaint-handling and resolution, including out-of-court redress or alternative-dispute resolution processes? To what extent should these requirements align with similar requirements in the DSA? What current practices could be regarded as best practice? How frequently should VSPS providers be obliged to report to the Commission on their complaint handling systems and what should those reports contain? Should there be a maximum time-period for VSPS providers to handle user complaints and if so, what should that period be?

Providers should:²⁹

- Provide prominent, accessible and easy-to-use tools to help children and parents seek redress, including by highlighting how to use them during the sign-up/induction process and tailoring tools to the age of the child
- Provide children and parents access to expert advice to support their decision-making and help them understand their rights
- Have clear penalties applied fairly and consistently
- Offer opportunities to appeal decisions, and escalate unresolved appeals to expert third parties or regulators
- Provide response times that are appropriate to the seriousness of the report being made, including by responding immediately to children who appear to be in distress
- Provide children and parents with opportunities to correct a child's digital profile/footprint, with clear and accessible tools that match up to a child's data rights
- Inform children of action taken in redress processes by granting access to the status of their reports, communicating actions clearly and giving them the opportunity to provide feedback

Procedural aspects of complaint-handling and resolution, as well as reporting frequency such complaint handling systems and content requirements related to such reports,

²⁷ Tick to Agree, 5Rights Foundation, [link](#)

²⁸ IEEE Standard for an Age Appropriate Digital Services Framework Based on the 5Rights Principles for Children, IEEE Consumer Technology Society, [link](#)

²⁹ IEEE Standard for an Age Appropriate Digital Services Framework Based on the 5Rights Principles for Children, IEEE Consumer Technology Society, [link](#)

should be aligned with requirements and procedures established under the DSA, so as to avoid duplications and ease compliance for VSPS providers.

5.3.2/3 Risk assessments and Safety by Design

Q18 What approach do you think the Code should take to risk assessments and safety by design? Are there any examples you can point us towards which you consider to be best practice?

VSPS providers should assess the risks outlined in the 4Cs framework³⁰ presented by each feature of the product or service to reveal known harms, potential risks and unintended consequences. VSPS providers should consider the potential for their own features to negatively impact children, with attention paid to how features may be experienced differently when encountered in combination, and how they might impact different groups of children. At the end of this process, providers will be able to identify elements or features that may need to be disabled, redesigned or carry warnings and/or other mitigation measures to keep children safe. It will also allow providers to make positive changes that deliver enhanced, creative and age-appropriate experiences. VSPS providers should consider both the likelihood of harm occurring and the severity of harm when it does occur. The likelihood of a child encountering harm can be measured by, among other methods, peer-reviewed academic research, internal research, A/B testing and data from public bodies. VSPS providers should make use of child development experts, official advice from public health authorities and the testimony of children themselves when measuring the severity of harm. VSPS providers should recognise how the risks created by individual features can increase when they are used in combination with other features. VSPS providers should consider the complexity of risk when assessing their service and features. Approaches to risk assessment and mitigation should be based on “the best available information and scientific insights.” VSPS must mitigate and manage the risks and design their services with children’s safety in mind.

While conducting their risk assessments, VSPS providers should consider many factors:

Using features in combination: VSPS providers should recognise how the risks created by individual features can increase when they are used in combination with other features. For instance, a service which makes use of algorithmic friend recommendations that recommend child accounts to adults and make a child’s location discoverable by other users would make the potential for grooming and sharing CSAM more likely.

Misuse of features: VSPS providers should account for how their features might be misused by actors with malign intent. Such risks may arise through:

- Inauthentic use of the service, such as the creation of fake accounts
- The use of bots or deceptive use of the service
- Other automated or partially automated behaviours
- Coordinated manipulation and use of their services
- Systemic infringement of their terms of service

VSPS providers should pay particular attention to how their services, and any use of algorithmic amplification, may contribute to these systemic risks.

Risks over time: Certain risks may expose children to low levels of immediate harm but increase in severity over time. A single notification may momentarily distract a child, for instance, but over time may have a more serious impact on their sleep, schoolwork and

³⁰ The 4Cs: Classifying Online Risk to Children. (CO:RE Short Report Series on Key Topics). Hamburg: Leibniz-Institut für Medienforschung | Hans-Bredow-Institut (HBI); CO:RE - Children Online: Research and Evidence, Livingstone, S., & Stoilova, M. (2021), [link](#)

ability to concentrate. VSPS providers should consider both the immediate and longer-term impacts their features have on children. Children can be exposed to risks over time in many ways, including:

1. Isolated exposure to risks that cause immediate harm, such as seeing a violent, sexual or otherwise developmentally inappropriate content
2. Cumulative passive exposure to risks over time, such as seeing the same narrow ideals of beauty consistently promoted in newsfeeds or timelines
3. Cumulative active engagement with risks, such as participating in pro-anorexia or self-harm groups

Similarly, the impact of harm can be either:

- Immediate or delayed – whether the impact of the experience occurred immediately after exposure or manifested at a later point
- Direct or indirect – whether the impact of the hazard occurred through direct exposure to the child who was harmed or indirectly through exposure

Children's vulnerability to harm: VSPS providers should consider the particular vulnerabilities of different children, at different development stages. For example, some children may be sharing devices with older siblings, and a six-year-old will require different protection measures than a 16-year-old. Other factors to consider include, but are not limited to: lack of digital access; low self-esteem; cognitive development issues; Mental or physical illness; having previously been a victim or perpetrator of conduct harms; lack of parental or guardian support; socio-economic deprivation; family difficulties; disability; educational disadvantage.

Risks to groups and society: As well as presenting risks to individual children, products and services might also pose risks to certain groups and wider society. For instance, the abuse of women online may have a chilling effect on girls' self-expression.³¹ Automated decision-making, has been shown to discriminate against certain groups when trained on poor datasets, such as those trained only on adults or on a particular gender or race.³² Young people's trust in democratic processes and institutions is undermined if services that encourage virality also allow disinformation to spread widely. VSPS providers have a responsibility to consider and address these collective and societal risks that impact children, as well as risks experienced by children as individuals.

5.3.5 Harmful feeds and recommender systems

Q20 What approach do you think we should take in the Code to address feeds which cause harm because of the aggregate impact of the content they provide access to? Are there current practices which you consider to be best practice in this regard?

While recommender systems can play an important role in helping children navigate the online world, to refine the masses of content available in a way that is supportive and can diversify their information ecosystem, they are more often configured to meet commercial goals. 5Rights Foundation's Risky by Design case study identified nine features which make use of these automated decision-making processes in ways that can lead to harm:

- Advertising: children should not be targeted and it should be clear when content is sponsored or paid-for. The prohibition of targeted advertising for children is established in the DSA.
- People also liked: children should not be compared with adults for 'people also liked...' features as inappropriate or dangerous material can be promoted to children.

³¹ No space for violence against women and girls in the digital world, Council of Europe, [link](#)

³² AI bias: exploring discriminatory algorithmic decision-making models and the application of possible machine-centric solutions adapted from the pharmaceutical industry, Belenguer, L., *AI Ethics* 2, 771–787 (2022), [link](#)

- Improved experiences: Children should have accessible options to prioritise the type of posts they want to see or turn off personalisation altogether.
- Filter bubbles: providers shall ensure that children are recommended a diverse range of content, to expand their horizons and burst filter bubbles.
- Ranking: VSPS should place less importance on 'popularity' and 'performance' when ranking recommendations, to broaden the variety and strengthen the veracity of information children are able to access.
- Autocomplete: VSPS should not recommend offensive or age-inappropriate suggestions for autocomplete.
- Recommending ideals: VSPS should assess the impact of algorithms used in recommendation systems, considering the objectives, data inputs, the rules which weight information with more or less importance, and the intended and unintended outcomes.
- Shadow banning: VSPS should provide information on how content has been ranked, showing the data and algorithms used to arrive at a decision.
- Friend/follower suggestion: VSPS should restrict adults from seeing children's accounts in friend or follower recommendations, and should not show young people's content to adults as suggested content.

A lack of transparency around the design and operation of recommendation systems is a major obstacle to addressing the risks they create. Providers should follow best practices when designing and operating recommendation systems. With greater transparency and effective oversight – and privacy-preserving age assurance that gives children age-appropriate experience online - those designing digital products and services can reduce the risks that recommendation systems pose to young people. Certain providers of online services will already be pursuing modifications to how their recommender systems are designed or operated in order to comply with the DSA. Tik Tok, for instance, has declared that it will allow users to turn off personalised recommendations for videos. Following a safety by design approach, for minors such recommendations should be off by default. As mentioned in previous replies, risks deriving from aggregate impact of content should be always carefully considered and mitigated.



Age Verification Providers Association

Laura Forsythe
Coimisiún na Meán,
2 – 5 Warrington Place,
Dublin D02 XP29

VSPSregulation@cnam.ie

September 4, 2023

Dear Ms Forsythe,

The Age Verification Providers Association is the global trade body for providers of online age assurance technology, including both age verification and age estimation solutions. We are pleased to respond to this consultation and look forward to working with the Commission as it establishes the regulatory regime, not only for Ireland but for a number of major platforms which operate in the European Union and fall under your jurisdiction.

Question 1: What do you think our main priorities and objectives should be in the first binding Online Safety Code for VSPS? What are the main online harms you would like to see it address and why?

Age assurance is the foundation for all aspects online safety for children because without it, platforms have no idea which of their users are children. They cannot offer children the higher level of protection they deserve, and are already given in the real world, if they cannot be differentiated from adult users.

So, while the consultation focuses on the implementation of the Audio-Visual Media Services Directive, which allows for both age assurance and parental controls, this is an opportunity to establish a standards-based, privacy-preserving approach to re-usable and interoperable age assurance, enabling the Commission to fulfil its wider responsibilities towards online safety in general, particularly where relying on parental discretion, which is itself dependent on the awareness, capability and will of parents to implement controls, is insufficient to provide a comprehensive level of protection to ALL children, as appropriate to their age.

Question 2: What types of online harms do you think should attract the most stringent risk mitigation measures by VSPS? How could we evaluate the impact of different types of harms e.g., severity, speed at which harm may be caused? Is there a way of classifying harmful content that you consider it would be useful for us to use?

Age assurance should be applied in proportion to the risk of harm. It is for policymakers and yourselves to determine what level should be applied, and we can deliver a full range of methods for only age checks which apply the mildest of assurance that a user is around a certain age, to proof beyond reasonable doubt that they are the age they claim to be.



Age Verification Providers Association

We recommend the Commission considers specific standardised *levels of age assurance*, such as those referenced in the forthcoming international standard IEEE P2089.1 and outlined in recent [research](#) commissioned by the UK Digital Regulation Cooperation Forum ([DRCF](#)) from the [Age Check Certification Scheme](#) for each use-case. This will lead to the Commission regulating on the basis of outputs - the overall level of accuracy and reliability of age assurance solutions - rather than being drawn into approving particular methods of age assurance. That allows for innovation and can mean that the level of rigour improves as technology improves, rather than merely achieving the level required by regulator and ceasing further efforts to improve the accuracy and reliability of age checks.

So, for example, a standard level of age assurance may be considered sufficient to create a social media account at the age of 16 without parental consent, but to use its dating facility may require a strict or enhanced level of check to ensure the user is 18 or older before they meet a stranger in real life.

Question 3: Do you have reports, academic studies or other relevant independent research that would support your views? If you do, please share them with us with links to relevant reports, studies or research.

We would like to commend to the Commission the extensive research into age assurance carried out as part of the [euCONSENT](#) project by Professors Simone van der Hof, Sonia Livingstone and Abhilash Nair. This can be found [here](#) and included a rapid evidence review, analysis of the EU legal framework, methods of AVMSD and GDPR Compliance and methods for obtaining parental consent under Article 8 of GDPR.

The UK Information Commissioner's Office has been working with Ofcom to find out more about some aspects of age assurance, to inform future work. It has so far published three important papers:

Measurement of Age Assurance Technologies - DCRF research commissioned by the ICO and Ofcom

"The ICO commissioned a technical study of measures to assess the effectiveness of age assurance techniques and to understand the potential for consistency, comparability and standardisation of measurement."

[Measurement of age assurance technologies - part 1](#)

"Age assurance is a complex area with technology developing rapidly. Under the Digital Regulation Co-operation Forum (DRCF), Ofcom and the ICO jointly commissioned research to explore ways of measuring the accuracy levels achievable by different age assurance solutions. It encourages further reflection on how to measure the overall effectiveness of age assurance methods, in addition to their technical accuracy. We will continue to work together to ensure regulatory alignment in this area."

[Measurement of age assurance technologies – part 2 \(22 August 2023\)](#)

Families' attitudes to age assurance

"Under the Digital Regulation Co-operation Forum (DRCF), Ofcom and the ICO jointly commissioned research to explore the attitudes of children and parents/carers to age assurance methods and where they see the balance of trade-offs between considerations such as privacy, online safety and ease of use."



Age Verification Providers Association

The research involved researchers speaking to parents and children individually, as well as focus groups where children and parents were able to discuss age assurance methods in more depth.”

[Families’ attitudes towards age assurance: research commissioned by the ICO and Ofcom](#)

Question 4: What approach do you think we should take to the level of detail in the Code? What role could non-binding guidance play in supplementing the Code?

In terms of age assurance, this is a rapidly developing technology and it would stifle innovation to be too prescriptive about which methods should be adopted, so we favour your Option 3, a mixed approach where the Commission is clear about the objectives and the outcomes it expects to be delivered by online services when they check the age of their users, but leaves more discretion about how this is achieved.

In answer to question 22 we state our support for co-regulation. The Commission should position itself to reference international standards whereby services which meet those standards can consider themselves to be compliant with the Commission’s requirements in that field. While this should not provide immunity from enforcement action, it should be acknowledged as effective due diligence by services when they seek to comply, and considered in mitigation when things go wrong.

Question 10a: What requirements should the Code include about age verification and age assurance?

We support a proportionate approach to age assurance, with the Commission stipulating the level of age assurance required for the most common use-cases, based on the latest international standards.

The Commission should be technologically neutral in specifying methods of age assurance. The most suitable method will depend upon the nature of the service, the availability of data on which to base age checks, and the level of age assurance required. Consumer choice is important, as it promotes accessibility by offering a wide range of methods.

It is an error to assume that longstanding age verification methods, such as the use of government ID or referencing authoritative databases such as credit reference agencies, generally provide a higher level of age assurance than age estimation methods which, for example, need not authenticate that the user is the rightful owner if evidence being offered for their age is biometric. Each method must be independently tested and certified before conclusions are drawn about its accuracy and overall reliability.

The Commission should support reusability and interoperability to promote a user-friendly solution to age assurance. If users cannot re-use an age check and use it across multiple services, they will quickly become frustrated with the process. The large-scale trial of interoperable, reusable age checks successfully delivered by [euCONSENT](#) demonstrates this is quite feasible.

The consultation states: ***“There is a proposal for a European Digital Identity that would provide EU citizens with a means of proving their age without disclosing any other personal data, but this is not yet in place”***.

CAUTION: The Commission should not assume that the eIDAS Wallet will solve the challenge of online age assurance. We have engaged with DG Connect on this question in a number of forums and it is not apparent how the wallet will facilitate user-friendly online age assurance. The security level of the wallet is designed to be ‘High’ which means there will be a rigorous procedure required each time the wallet is used – similar to logging into your online banking. This may be tolerable

THE AVPA LIMITED OPERATING AS THE AGE VERIFICATION PROVIDERS ASSOCIATION (COMPANY NO 11961982)

REGISTERED ADDRESS: 557B WANDSWORTH ROAD, LONDON, SW8 3JD

GENERAL ENQUIRIES: AVPA@AVPASSOCIATION.COM MEDIA ENQUIRIES: PRESS@AVPASSOCIATION.COM



Age Verification Providers Association

when a user is opening a new pension account or buying a house, but will soon become very frustrating when surfing the web, accessing perhaps 20 websites an hour which each require age assurance to ensure only age-appropriate content is shown. We believe an additional layer will be required in the technological 'stack' to facilitate making the Internet "age-aware". **Giving consent to every data controller and processor for each website you browse to retrieve your age attributes from your eIDAS wallet is not a viable option under its current design ambitions.** We note also that presently eIDAS is used mostly by adults and older teenagers; it may be many years before 7, 13, or even 16 year-old ubiquitously own a European digital identity. There is also, in some use-cases particularly, a question as to how comfortable a citizen would be to use any ID directly issued by their government.

(By the same token, citizens may also not wish to be forced to rely on Very Large Online Platforms which already have the opportunity to gather enormous amounts of personal data, to surf the web in a compliant manner).

Question 10b: What sort of content should be shown by default to users who are logged out or in private browsing mode and whose age cannot be verified or assured?

This question confines itself to content, but of course online harms arise also from contact, conduct and commerce (the 4 C's).

It would be illegal under Article 8 of GDPR and against the advice of the "Fundamentals" guidance of the Data Protection Commissioner for any online service to process personal data on the basis of consent if the user is under 16 in Ireland (and various ages across Member States from 13-16), so that should first be considered.

There are then wider questions around age-appropriate content, including but not limited to video content, and other harmful functionality, including algorithmic content selection and targeting of advertisements. So, the only safe approach is for services to be designed to be safe for users of all ages, unless they apply proportionate age-assurance.



Age Verification Providers Association

Question 10c: What evidence is there about the effectiveness of age estimation techniques?

We would signpost the Commission to the following selection of evidence from some of our members already in the public domain as indicative of the latest levels of accuracy for facial image and voiceprint analysis.

Yoti

Mean Absolute Error by age band

YOTI Yoti facial age estimation accuracy		Mean estimation error in years split by gender, skin tone and age band								
Gender	Female				Male				All	
Skintone	Tone 1	Tone 2	Tone 3	All	Tone 1	Tone 2	Tone 3	All		
6-12	1.3	1.4	1.6	1.4	1.2	1.4	1.3	1.3	1.3	
13-17	1.3	1.5	1.6	1.5	1.0	1.3	1.5	1.3	1.4	
18-24	2.4	2.5	2.6	2.5	2.0	2.0	2.2	2.0	2.3	
25-70	2.9	3.3	4.6	3.6	2.6	3.3	3.6	3.2	3.4	
6-70	2.5	2.9	3.8	3.1	2.3	2.8	3.0	2.7	2.9	

<https://www.yoti.com/blog/yoti-age-estimation-white-paper/>

Privately SA

Independent Testing: In March 2023, ACCS, a UKAS accredited conformity assessment body, examined VoiceAssure and noted the following:

- A 100% accuracy in predicting 13–14-year-olds as minors.
- A 100% accuracy in predicting 26–27-year-olds as adults.

<https://medium.com/@onuryrten/voiceassure-a-robust-privacy-first-voice-based-age-estimation-technology-f9c9d6c8340d>

	Female			Male		
Fitzpatrick Scale (Skin Tone)	I&II	III&IV	V&VI	I&II	III&IV	V&VI
Age						
0-12	0.8	0.89	0.97	0.77	0.72	0.87
13-17	0.68	1.2	0.94	0.97	1.03	1.13
18-24	1.77	1.92	2.29	1.05	1.95	1.97
25-65	2.96	2.38	2.43	2.92	2.94	2.46

<https://medium.com/@onuryrten/faceassure-developing-a-robust-privacy-first-face-based-age-estimation-technology-2d1e5e096230>



Age Verification Providers Association

VerifyMyAge

VerifyMyAge offer age estimation based a range of methods and their accuracy can be found here, which is also an example of the certification available to providers from the UK Accreditation Services approved Conformity Assessment Body, the Age Check Certification Scheme.

<https://www.accscheme.com/media/2b4ffso1/schedule-of-certification-pas-1296-updated-verison-3.pdf>

Of note is their Fully Qualified Domain Addresses (FQDA) method which is analysis of the usage of a user's email address, providing another, highly effective method of age estimation without the use of biometric information which some users prefer not to share.

There is also a company in Portugal developing age estimation based on how users play computer games, as an example of how there is constant innovation in this field.

Question 10d: What current practices do you regard as best practice?

While we do not argue for wholly independent age assurance carried out only by third parties, the Commission should consider both the question of confidentiality and expertise.

- When the use-case relates to a sensitive field such as adult content, users may prefer to share, to the extent it is necessary (and that might be just sharing a voice sample of course) with an independent third party age verification provider who is subject to close supervision by the DPC or its equivalents, and has an existential interest in delivering privacy-by-design and data-minimisation.
- Where accuracy is important, again a specialist provider may deliver better results than a platform that invests in its own solution amongst many other priorities.

It may be that the Commission accepts lower levels of age assurance can be achieved in-house, but seeks objective audit and certification for higher levels of age assurance, which may encourage the use of specialists without excluding internal systems altogether.

Question 10e: Where accounts are not age verified should default privacy settings be used, should content default to universal content and should contact by others be more limited?

See Question 10 above.

Question 12: What requirements should the Code have in relation to parental control features? How can we ensure that VSPS providers introduce the mechanism in a user-friendly and transparent way? Can you point to any existing example of best practice in this area? Should parental controls be 'turned-on' by default for accounts of minors or where age is not verified?

It is essential to understand that parental controls deliver a substantively different policy objective from online age assurance. Parental controls allow parents to exercise their discretion effectively in determining what their children can do online. This does not enforce decisions made by the legislature or regulators as to what content is unsuitable for children. Parents must also be aware of these controls, be capable of turning them on and make a decision to do so. It will depend upon the particular use case as to whether it is more appropriate to require parental controls or online age assurance.

Question 14: How should we ask VSPS providers to address online harms in their terms and conditions in the Code, including the harms addressed under Article 28b? How should key aspects



Age Verification Providers Association

of terms and conditions be brought to users' attention? What examples are there of best practice in relation to terms and conditions including content moderation policies and guidelines?

Terms and conditions should address age assurance. They should be clear how age assurance is achieved and by whom. If personal data used for age assurance is re-used for the purposes, such as targeted marketing or even birthday promotions, this should be made explicit.

Question 16: What requirements should the Code include about procedures for complaint-handling and resolution, including out-of court redress or alternative-dispute resolution processes? To what extent should these requirements align with similar requirements in the DSA? What current practices could be regarded as best practice? How frequently should VSPS providers be obliged to report to the Commission on their complaint handling systems and what should those reports contain? Should there be a maximum time-period for VSPS providers to handle user complaints and if so, what should that period be?

In terms of age assurance, it is important that users can seek redress if their age has been assessed incorrectly. Generally, providers offer alternative methods if the first method chosen does not deliver an accurate result. This is important to deliver accessibility and avoid discrimination.

Where all options are exhausted, platforms should ensure users can contact their AV provider to seek corrective action.

Question 17: What approach do you think the Code should take to ensuring that the safety measures we ask VSPS providers to take are accessible to people with disabilities?

Accessibility is a priority for the AV sector as a whole. It is delivered through the wide range of methods available to prove your age online. Re-usable age checks are of particular importance to those who may need assistance in completing such a check, as the process need not be repeated as frequently. Interoperability takes this one stage further, so we commend to the Commission the work of [euCONSENT ASBL](#) to which our members contribute.

Question 18: What approach do you think the Code should take to risk assessments and safety by design? Are there any examples you can point us towards which you consider to be best practice?

Privacy-by-design and data minimisation are already required under GDPR but the Commission should emphasise this requirement for age assurance solutions to reassure the public about data security and privacy. Our [Code of Conduct](#) seeks to export European best practices to our members globally and is a requirement of full membership of our Association.

Question 19: How do you think that cooperation with other regulators and bodies can help us to implement the Code for VSPS?

We have worked closely with the UK ICO and Ofcom and we recommend the research they have commissioned into age assurance. As a world-wide web, there is a strong case for aligning regulations, or at least shape and dimensions of it, globally. That is why we champion international standards from BSI, IEEE and ISO.

Question 22: What compliance monitoring and reporting arrangements should we include in the Code?



Age Verification Providers Association

We strongly support the concept of co-regulation. This leverages the work of private sector conformity assessment bodies to audit and certify platforms to international standards. The Commission can approve specific Certification Schemes and incentivise platforms to subject themselves to audit in return for more leniency if they are subsequently found to have breached the code.

There is already an international standard for online age verification – BSI PAS 1296:2018 and the [Age Check Certification Scheme](#) has been approved by the UK Accreditation Service to issue certification of compliance with that standard. Updated standards are due from the IEEE this year and ISO in 2024/25.

Question 23: Should the Code have a transition period or transition periods for specific issues? Which areas touched on in this Call for Inputs may VSPS providers require time to transition the most? What time frame would be reasonable for a transition period?

Implementing online age assurance is not a lengthy process, even for very large online platforms. Many of those have already implemented it for parts of the service and in other jurisdictions. Smaller services can usually take advantage of API or plug-ins to the major content and e-commerce platforms offered by AV providers.

The UK Government set three months as the standard when it was preparing to implement age verification in 2019, and we are confident that all those within scope of the regulation would be able to put a solution in place within that notice period, given that the underlying legal requirements for age assurance are already well known. Adult sites which chose to implement age verification when required to do so by the French regulator managed this in ten days.

In conclusion, we stand ready to assist the Commission in its duties, and look forward to working in partnership to make the Internet a safer place for children.

Yours sincerely,

Iain Corby

Executive Director

Age Verification Providers Association



**Belong To Submission to Coimisiún na Meán on
Developing Ireland’s First Binding Online Safety Code for
Video-Sharing Platform Services**

4th September 2023

Contact: Neasa Candon (neasa@belongto.org), Moninne Griffith
(moninne@belongto.org)

Table of Contents

Introduction.....	2
Belong To’s Online Safety Work.....	3
Research Background: LGBTQ+ Youth and Social Media.....	4
Question 1: Main priorities, objectives and online harms	7
Question 2: Stringent mitigation, evaluation and classification.....	9
Question 3: Reports, academic studies and relevant independent research.....	10
Question 4: Code detail and non-binding guidance.....	12
Question 7: Content connected to video content.....	12
Question 9: Flagging mechanism, transparency and user-friendly design.....	13
Question 10: Age verification and age assurance.....	15
Question 11: Content rating.....	15
Question 12: Parental control features.....	16
Question 13: Media literacy measures and tools.....	17
Question 14: Terms and conditions.....	17
Question 15: Content moderation.....	18
Question 16: Complaint-handling, resolution and reporting.....	19

Question 18: Risk assessments and safety by design.....	19
Question 23: Transition periods and timeframes.....	20

Introduction

Belong To, LGBTQ+ Youth Ireland is a national organisation supporting lesbian, gay, bisexual, transgender, and queer (LGBTQ+) young people. Since 2003, Belong To has worked with LGBTQ+ youth to create a world where they are equal, safe and thriving in the diversity of their identities and experiences.

The organisation advocates and campaigns with and on behalf of LGBTQ+ young people and offers specialised LGBTQ+ youth services in Dublin (including crisis counselling with Pieta) and supports a network of LGBTQ+ youth groups across Ireland. Belong To also supports educators and other professionals working with LGBTQ+ youth with training, capacity building and policy development.

We strongly welcome the opportunity to contribute to the meaningful work of Coimisiún na Meán on developing Ireland's first binding Online Safety Code for video-sharing platform services. As detailed in the following sections, LGBTQ+ youth occupy a relatively unique position in relation to online safety. While LGBTQ+ young people are particularly vulnerable to online harms, including anti-LGBTQ+ hate speech and cyberbullying, online spaces are also an important source of information, support and community for LGBTQ+ youth.

Along with addressing questions raised in the Call for Inputs, our submission also raises the need for a comprehensive approach to three key areas:

- Addressing online hate speech and content that incites violence;
- The need for protections including parental controls and age verification; and
- The balance of LGBTQ+ young people's position as rights-holders.

Belong To looks forward to continued engagement with Coimisiún na Meán on the development of this code, and the future work of the Commission in relation to youth participation, complaints handling, and accountability mechanisms for online platforms.

Belong To's Online Safety Work

Online safety is a key strategic priority for Belong To. In relation to policy, Belong To is a member of the Children's Rights Alliance Online Safety Advisory Group, and engaged extensively in the development of the Online Safety Media Regulation Act. The importance of digital literacy, and empowering young people with the information needed to navigate online spaces safely, were key elements of our submission to the National Council on Curriculum and Assessment (NCCA) as part of the review of the Social Personal and Health Education (SPHE) curriculum for Junior Cycle students.¹ We were pleased to see a number of recommendations relating to online safety, digital literacy and the rights of young people online included in the final curriculum.

Belong To has also been proactive in developing relationships with social media platforms, and VSPS platforms. Funded by the Google.org 2019 Impact Challenge on Safety, Belong To has run the annual 'It's Our Social Media' campaign since 2022. 'It's Our Social Media' is a digital media campaign combatting online hate speech experienced by LGBTQ+ youth, while empowering young people to take back social media, protect themselves online, and to hold social media companies accountable as we work to make spaces safe for users. The campaign features a range of digital assets, including a short-form animation and a hero video of young LGBTQ+ people sharing their thoughts on social media and how we can create safe spaces for LGBTQ+ youth online. We have two key campaign slogans; #FeedTheGood and #BlockTheBad, both of which helped us prompt users to join in the conversation online and take action.

Another key component of this campaign was our microsite, itsoursocialmedia.com, which acted as an online hub that housed resources on how to stay safe online, digital self-care tips and much more. The microsite also featured an online poll to gather users' thoughts on online safety. The campaign ran across Facebook, Instagram, Twitter, Snapchat, and TikTok, generating over 11 million impressions, 3.3 million video views and 35,700 link clicks in 2022.

¹ Belong To (2022) 'Draft Specification for Junior Cycle SPHE – NCCA Consultation'. [Available here](#).

Research Background: LGBTQ+ Youth and Social Media

Online Harms and LGBTQ+ Youth

Internationally, LGBTQ+ youth are found to be more likely to experience bullying or harassment online than their non-LGBTQ+ peers, and less likely to feel safe while using social media.² Research shows that anti-LGBTQ+ online hate leads to LGBTQ+ youth feeling inferior and shameful about their identity, therefore developing an internalised sense of blame for the hateful content they witnessed.³ In response, LGBTQ+ young people were found to have developed the long-term coping strategies of isolating themselves socially, or repressing the visibility of their LGBTQ+ identity in public and community spaces.

Earlier this year, Belong To released findings relating to the experiences of LGBTQ+ young people living in Ireland and their social media use.⁴ A shocking 87% of LGBTQ+ youth had seen or experienced anti-LGBTQ+ hate and harassment on social media in the past year. 65% of LGBTQ+ young people surveyed had reported this content to a social media platform. Among young people who reported this content, only 21% saw action from the relevant social media platform; anti-LGBTQ+ content was removed in 12% of cases, 4% saw the offending user temporarily suspended, and 5% of reports resulted in the offending account being banned. The remaining 79% of LGBTQ+ young people were either informed that no violation of community guidelines was found or received no response from the platform.

Published in 2016, the *LGBT Ireland Report* found that 23% of LGBTQ+ participants reported having hurtful things written about them on social media. This was proportionately higher among trans people, at 34%, and among LGBTQ+ participants aged 14-25, at 32%.

An increase in the far-right movement globally has mapped a wide-scale increase in anti-LGBTQ+ hate, harassment and discrimination, both online and offline. Social

² GLSEN (2013) *Out Online: The Experiences of Lesbian, Gay, Bisexual and Transgender Youth on the Internet*. GLSEN: New York. [Available here](#).

³ Keighley, R. (2022) 'Hate Hurts: Exploring the Impact of Online Hate on LGBTQ+ Young People', *Women & Criminal Justice*, 32:1-2, 29-48. [Available here](#).

⁴ Pizmony-Levy, O. (2022) *The 2022 Irish School Climate Survey*. Research Report. Global Observatory of LGBTQ+ Education and Advocacy. Dublin and New York: Belong To and Teachers College, Columbia University. [Available here](#).

media algorithms have served to facilitate and promote this proliferation of hateful content and disinformation.

As documented by organisations such as Hate Aide, social media platforms have allowed for the convergence of far-right, right-wing, radical right, religious extremist, anti-LGBTQ+ and Covid-sceptic actors, fuelled by an algorithmic business model that understands the mass engagement with and dissemination of this content as profitable.⁵ This has increasingly resulted in real-world, hate-motivated violence, particularly against LGBTQ+ people.⁶

The European Digital Media Observatory (EDMO), an international organisation that seeks to analyse disinformation, reported in May of this year that “mis- and disinformation targeting the LGBTQ+ community is one of the most present and consistent in the European Union”.⁷ Research conducted in 2021 found that LGBTQ+ people experience 50% more online hate and harassment than any other minority group.⁸ In Ireland, as relates to the LGBTQ+ community, this has primarily manifested as disruptive ‘protests’ opposing the availability of books which represent LGBTQ+ experiences and identities in public libraries.⁹

Benefits of Online Spaces for the LGBTQ+ Community

Despite the above outlined harms, it is important to highlight the importance of social media and online spaces for LGBTQ+ young people, and to ensure their continued access to content that is informative, entertaining and inclusive.

International research shows that LGBTQ+ young people use social media at much higher rates than non-LGBTQ+ youth, often to seek community and to look for the safe spaces and information they may not have access to in real life.¹⁰ In an Irish context,

⁵ Hate Aid (2023) ‘Small changes – big effect: how hate on the internet can be reduced’. [Available here](#).

⁶ Squirrel, T. and Davey, J. (2023) *A Year of Hate: Understanding Threats and Harassment Targeting Drag Shows and the LGBTQ+ Community*. Institute of Strategic Dialogue: London. [Available here](#).

⁷ Panizio, E. and Canetta, T. (2023) ‘Rights in the time of conspiracies and fake news: disinformation against LGBTQ+ in the EU’. European Digital Media Observatory: Italy. [Available here](#).

⁸ ADL Centre for Technology & Society (2021) *Online Hate and Harassment: The American Experience*. ADL: New York. [Available here](#).

⁹ Fitzgerald, C. (2023) ‘Explainer: Why is the far-right targeting Ireland's libraries?’, *The Journal.ie*. [Available here](#).

¹⁰ Steinke, J. Root-Bowman, M. Estabrook, S. Levine, D. Kantor, L. (2017) ‘Meeting the Needs of Sexual and Gender Minority Youth: Formative Research on Potential Digital Health Interventions’, *Journal of Adolescent Health* 60(5). [Available here](#).

this source of community and support is particularly important for LGBTQ+ youth, 56% of whom live in home environments that are not supportive of their LGBTQ+ identity.¹¹

As part of the *LGBT Ireland Report*, participants were asked about their experiences of coming out, and finding support and information relating to this.¹² The internet, social media and traditional media were identified as the most significant practical elements in helping participants to come out. Social media was named as useful in finding out about LGBTQ+ identities, getting advice on approaches to coming out, and exploring one's own identity. Relating to this Call for Inputs in particular, several participants named accessing others' experiences of identifying as LGBTQ+ and coming out through YouTube videos as an important source of hope, inspiration and advice. One participant shared:

“Hearing people’s stories and experiences on YouTube was invaluable to me. YouTube was also extremely helpful to see people living their lives happily while out of the closet. (Gay male, 19)”

¹¹ Belong To (2021) *LGBTI+ Life in Lockdown: One Year Later*. Dublin: Belong To. [Available here](#).

¹² Higgins A. et al. (2016) *The LGBTIreland Report: national study of the mental health and wellbeing of lesbian, gay, bisexual, transgender and intersex people in Ireland*. Dublin: GLEN and Belong To. [Available here](#).

Responses to Questions Posed in the Call for Inputs

Question 1: Main priorities, objectives and online harms

What do you think our main priorities and objectives should be in the first binding Online Safety Code for VSPS?

The main priority in the first binding Online Safety Code for VSPS should be the protection of children and young people online.

In terms of objectives, the following are required:

- A robust response to hate speech and extreme material, as aligned with the incoming Criminal Justice (Incitement to Violence or Hatred and Hate Offences Bill);
- Addressing the issue of algorithmic promotion of hateful and extreme content; and
- Clear requirements for social media platforms relating to reporting, the platform's response, and community guidelines.

The four areas set out in Article 28b of the Audio- Visual Media Services Regulation need to be addressed by the Online Safety Code, namely:

1. Content that might impair the physical, mental or moral development of minors. This includes content that may be inappropriate for children, such as pornography.
2. Content that incites violence or hatred against a group of persons or a member of a group based on any of the grounds referred to in Article 21 of the European Charter of Fundamental Rights. These grounds include sex, race, colour, ethnic or social origin, genetic features, language, religion or belief, political or any other opinion, membership of a national minority, property, birth, disability, age or sexual orientation.
3. Content the dissemination of which constitutes a criminal offence under EU law.
4. Certain commercial communications that would not be permitted on broadcast or video-on-demand services. Commercial communications include advertising, sponsorship and product placement.

The categories of harm set out in the Broadcasting Act 2009, as amended by the Online Safety and Media Regulation Act 2022, also should be addressed by the Online Safety Code:

- a) Harmful online content relating to 42 criminal offences under Irish law listed in Schedule 3 of the 2009 Act as amended;
- b) online content by which a person bullies or humiliates another person;
- c) online content by which a person promotes or encourages behaviour that characterises a feeding or eating disorder;
- d) online content by which a person promotes or encourages self-harm or suicide; and
- e) online content by which a person makes available knowledge of methods of self-harm or suicide.

What are the main online harms you would like to see it address and why?

The issue of hateful anti-LGBTQ+ content, and misinformation relating to the LGBTQ+ community and LGBTQ+ identities, is the main concern for Belong To in relation to the new Online Safety Code.

The European Digital Media Observatory (EDMO), an international organisation that seeks to analyse disinformation, reported in May of this year that “mis- and disinformation targeting the LGBTQ+ community is one of the most present and consistent in the European Union”.¹³ Research conducted in 2021 found that LGBTQ+ people experience 50% more online hate and harassment than any other minority group.¹⁴

Earlier this year, Belong To released findings relating to the experiences of LGBTQ+ young people living in Ireland and their social media use.¹⁵ A shocking 87% of LGBTQ+ youth had seen or experienced anti-LGBTQ+ hate and harassment on social media in the past year. 65% of LGBTQ+ young people surveyed had reported this

¹³ Panizio, E. and Canetta, T. (2023) ‘Rights in the time of conspiracies and fake news: disinformation against LGBTQ+ in the EU’. European Digital Media Observatory: Italy. [Available here](#).

¹⁴ ADL Centre for Technology & Society (2021) Online Hate and Harassment: The American Experience. ADL: New York. [Available here](#).

¹⁵ Pizmony-Levy, O. (2022) *The 2022 Irish School Climate Survey*. Research Report. Global Observatory of LGBTQ+ Education and Advocacy. Dublin and New York: Belong To and Teachers College, Columbia University. [Available here](#).

content to a social media platform. Among young people who reported this content, only 21% saw action from the relevant social media platform; anti-LGBTQ+ content was removed in 12% of cases, 4% saw the offending user temporarily suspended, and 5% of reports resulted in the offending account being banned. The remaining 79% of LGBTQ+ young people were either informed that no violation of community guidelines was found, or received no response from the platform.

Research shows that anti-LGBTQ+ online hate leads to LGBTQ+ youth feeling inferior and shameful about their identity; therefore developing an internalised sense of blame for the hateful content they witnessed.¹⁶ In response, LGBTQ+ young people were found to have developed the long-term coping strategies of isolating themselves socially, or repressing the visibility of their LGBTQ+ identity in public and community spaces.

The proliferation of conspiracy thinking, misinformation and disinformation relating to a range of communities and topics, including LGBTQ+ people, has increasingly resulted in real-world violence against LGBTQ+ individuals, spaces and events.¹⁷ In Ireland, this has also manifested as disruptive ‘protests’ in public libraries, opposing the availability of books which represent LGBTQ+ experiences and identities.¹⁸

Question 2: Stringent mitigation, evaluation and classification

What types of online harms do you think should attract the most stringent risk mitigation measures by VSPS?

Online harms which amount to criminal behaviour should attract the most stringent risk mitigation measures by VSPS.

How could we evaluate the impact of different types of harms e.g. severity, speed at which harm may be caused?

For a more detailed response to this question, we direct to the comprehensive submission compiled by the Children’s Rights Alliance.

¹⁶ Keighley, R. (2022) ‘Hate Hurts: Exploring the Impact of Online Hate on LGBTQ+ Young People’, *Women & Criminal Justice*, 32:1-2, 29-48. [Available here.](#)

¹⁷ Squirrell, T. and Davey, J. (2023) *A Year of Hate: Understanding Threats and Harassment Targeting Drag Shows and the LGBTQ+ Community*. Institute of Strategic Dialogue: London. [Available here.](#)

¹⁸ Fitzgerald, C. (2023) ‘Explainer: Why is the far-right targeting Ireland’s libraries?’, *The Journal.ie*. [Available here.](#)

Regarding severity, evaluation of different types of harms should keep in mind existing and incoming criminal laws relating to harm and abuse. In particular, this relates to child sexual abuse materials, intimate images and material that incites hatred.

Is there a way of classifying harmful content that you consider it would be useful for us to use?

The Council of Europe has recommended that ‘states should co-operate with a view to promoting standardisation of content classification and advisory labels among countries and across stakeholder groups to define what is appropriate and what is inappropriate for children’.¹⁹

In their submission, the Children’s Rights Alliance has provided a detailed break-down of three classification systems that could be considered, namely CO:RE 4Cs Classification; the Australian Classification System; and the UK Classification System.

Question 3: Reports, academic studies and relevant independent research

Do you have reports, academic studies or other relevant independent research that would support your views? If you do, please share them with us with links to relevant reports, studies or research.

Publications related to best practice in online safety cited in this submission include:

- 5Rights Foundation, Tick to Agree Age appropriate presentation of published terms September 202, 10-22. [Available here.](#)
- Council of Europe (2018) Guidelines to respect, protect and fulfil the rights of the child in the digital environment. Recommendation CM/Rec(2018)7 of the Committee of Ministers, 29, para 121. [Available here.](#)
- Hate Aid (2023) ‘Small changes – big effect: how hate on the internet can be reduced’. [Available here.](#)
- Information Commissioner’s Office, ‘Age Appropriate Design: A Code of Practice for Online Services’ 21. [Available here.](#)

¹⁹ Council of Europe, Guidelines to respect, protect and fulfil the rights of the child in the digital environment (2018) Recommendation CM/Rec(2018)7 of the Committee of Ministers, 29, para 121. [Available here.](#)

- UN Committee on the Rights of the Child (2021) General Comment no 25 on children's rights in relation to the digital environment, CRC/C/GC/25, para 73. [Available here.](#)

Publications related to LGBTQ+ youth and online harms more broadly cited in this submission include:

- ADL Centre for Technology & Society (2021) *Online Hate and Harassment: The American Experience*. ADL: New York. [Available here.](#)
- Bacchi, U. (2020) 'TikTok apologises for censoring LGBTQ+ content'. Reuters. [Available here.](#)
- Belong To (2022) 'Draft Specification for Junior Cycle SPHE – NCCA Consultation'. Dublin: Belong To. [Available here.](#)
- Belong To (2021) *LGBTI+ Life in Lockdown: One Year Later*. Dublin: Belong To. [Available here.](#)
- Fitzgerald, C. (2023) 'Explainer: Why is the far-right targeting Ireland's libraries?', *The Journal.ie*. [Available here.](#)
- GLSEN (2013) *Out Online: The Experiences of Lesbian, Gay, Bisexual and Transgender Youth on the Internet*. GLSEN: New York. [Available here.](#)
- Higgins A. et al. (2016) *The LGBTIreland Report: national study of the mental health and wellbeing of lesbian, gay, bisexual, transgender and intersex people in Ireland*. Dublin: GLEN and Belong To. [Available here.](#)
- Keighley, R. (2022) 'Hate Hurts: Exploring the Impact of Online Hate on LGBTQ+ Young People', *Women & Criminal Justice*, 32:1-2, 29-48. [Available here.](#)
- Panizio, E. and Canetta, T. (2023) 'Rights in the time of conspiracies and fake news: disinformation against LGBTQ+ in the EU'. European Digital Media Observatory: Italy. [Available here.](#)
- Pizmony-Levy, O. (2022) *The 2022 Irish School Climate Survey*. Research Report. Global Observatory of LGBTQ+ Education and Advocacy. Dublin and New York: Belong To and Teachers College, Columbia University. [Available here.](#)
- Squirrel, T. and Davey, J. (2023) *A Year of Hate: Understanding Threats and Harassment Targeting Drag Shows and the LGBTQ+ Community*. Institute of Strategic Dialogue: London. [Available here.](#)

- Steinke, J. Root-Bowman, M. Estabrook, S. Levine, D. Kantor, L. (2017) 'Meeting the Needs of Sexual and Gender Minority Youth: Formative Research on Potential Digital Health Interventions', *Journal of Adolescent Health* 60(5). [Available here](#).

Question 4: Code detail and non-binding guidance

What approach do you think we should take to the level of detail in the Code?

The Online Safety Code should take the form of a detailed prescriptive Code. As noted in the Call for Submissions, this would allow the Code to “specify details in the measures we expect VSPS providers to take to address online harms”.

Both protective and preventative measures should be included in the Code, namely prohibiting all forms of violence, exploitation and abuse; and including child-friendly mechanisms for consultation and participation, digital literacy supports for parents and carers, and effective remedies respectively.²⁰

Question 7: Content connected to video content

To what extent, if at all, should the Code require VSPS providers to take measures to address content connected to video content?

Consideration should be given to the Code requiring VSPS providers to take measures to address content connected to video content, such as captions and comments.

Earlier this year, Belong To released findings relating to the experiences of LGBTQ+ young people living in Ireland and their social media use.²¹ A shocking 87% of LGBTQ+ youth had seen or experienced anti-LGBTQ+ hate and harassment on social media in the past year. 65% of LGBTQ+ young people surveyed had reported this content to a social media platform. Among young people who reported this content, only 21% saw action from the relevant social media platform; anti-LGBTQ+ content

²⁰ UN Committee on the Rights of the Child, General Comment no 25 (2021) on children’s rights in relation to the digital environment, CRC/C/GC/25, para 73. [Available here](#).

²¹ Pizmony-Levy, O. (2022) *The 2022 Irish School Climate Survey*. Research Report. Global Observatory of LGBTQ+ Education and Advocacy. Dublin and New York: Belong To and Teachers College, Columbia University. Available at: https://www.belongto.org/wp-content/uploads/2022/11/2022-School-Climate-Survey_Full-Report.pdf

was removed in 12% of cases, 4% saw the offending user temporarily suspended, and 5% of reports resulted in the offending account being band. The remaining 79% of LGBTQ+ young people were either informed that no violation of community guidelines was found, or received no response from the platform.

In this research, community guidelines arose as a significant issue for young people attempting to report anti-LGBTQ+ content. It is vital that community guidelines are considered as part of this potential requirement, to ensure that, for example, harmful content posted as a comment in response to content that does not breach the code is treated as seriously as harmful video content. This is particularly important in relation to anti-LGBTQ+ bullying, and the fact that, in 2016, 34% of trans individuals, and 32% of LGBTQ+ people aged 14-25 living in Ireland reported having had hurtful things written about them on social media.²²

Question 9: Flagging mechanism, transparency and user-friendly design

How should we ask VSPS providers to introduce and design a flagging mechanism in the Code?

While a user flagging mechanism is important, it should not be a primary means relied upon to address harmful content, for two reasons. First, through the Code, VSPS and other social media site should be bound by a duty of care towards their users, meaning that the onus should be on social media platforms to address this harmful content before it reaches a critical mass of users.

Secondly, as stated previously, the process by which social media platforms respond to user reports has been found to be inconsistent. Research by Belong To shows that, of LGBTQ+ young people who reported anti-LGBTQ+ hate and harassment to social media platforms, only 21% saw action from the relevant platform; anti-LGBTQ+ content was removed in 12% of cases, 4% saw the offending user temporarily suspended, and 5% of reports resulted in the offending account being band. The

²² Higgins A. et al. (2016) *The LGBTIreland Report: national study of the mental health and wellbeing of lesbian, gay, bisexual, transgender and intersex people in Ireland*. Dublin: GLEN and Belong To. [Available here](#).

remaining 79% of LGBTQ+ young people were either informed that no violation of community guidelines was found, or received no response from the platform.²³

How can we ensure that VSPS providers introduce the mechanism in a user-friendly and transparent way?

Consideration should be given to consultation with children and young people when establishing what could be considered “user-friendly and transparent” in relation to flagging mechanisms.

Users should be able to track the progress of their report, and be provided with information as to a point of contact should the report take longer than a period of time specified by the Code to be addressed.

How should we ask VSP Providers to report the decisions they’ve made on content after it has been flagged?

Where content is deemed to have not breached community guidelines, VSP Providers should be required to provide clear reasoning for this upon request by the user. Users should be provided with a means of appealing such decisions to the VSP Provider.

To what extent should we align the Code with similar provisions on flagging in the DSA?

The DSA (Article 16) will require platforms to put in place a notification mechanism for illegal content and require them to process the notifications in a timely, diligent, non-arbitrary and objective manner. This should be integrated into the Code being developed.

Requiring users to determine whether they are flagging content under the DSA or the Code would place a significant burden on the user and could act as a deterrent to children and young people flagging illegal and harmful online content and, as such, would not be considered a user-friendly approach to integrating the DSA.

²³ Pizmony-Levy, O. (2022) *The 2022 Irish School Climate Survey*. Research Report. Global Observatory of LGBTQ+ Education and Advocacy. Dublin and New York: Belong To and Teachers College, Columbia University. [Available here](#).

Question 10:

What requirements should the Code include about age verification and age assurance?

For a more detailed response to this question, we direct to the comprehensive submission compiled by the Children's Rights Alliance.

Relating to age verification, there are a number of additional considerations to be taken in the case of LGBTQ+ young people. As stated previously, international research shows that LGBTQ+ young people use social media to seek community, and to look for the safe spaces and information they may not have access to in real life.²⁴ In an Irish context, this source of community and support is particularly important for LGBTQ+ youth, 56% of whom live in home environments that are not supportive of their LGBTQ+ identity.²⁵

As a result, consideration of the above should be given to age verification measures which require the input and/or consent of a parent, carer or guardian, balanced against rights enshrined under the UN Convention on the Rights of the Child to freedom of expression (article 13); freedom of thought, conscience and religion (article 14); freedom of association (article 15); and access to appropriate information (article 17).

Additionally, age verification measures should be cognisant of trans, non-binary and gender non-conforming young people, whose usernames and gender may not reflect that which is stated on official documentation.

Question 11: Content rating

What requirements should the Code have in relation to content rating?

For a more detailed response to this question, we direct to the comprehensive submission compiled by the Children's Rights Alliance.

Relating to LGBTQ+ youth, it is important that the Code require social media platforms follow best-practice guidelines in content rating, that are informed by LGBTQ+ identities and experiences. Experts in the area of online disinformation and

²⁴ Steinke, J. Root-Bowman, M. Estabrook, S. Levine, D. Kantor, L. (2017) 'Meeting the Needs of Sexual and Gender Minority Youth: Formative Research on Potential Digital Health Interventions', *Journal of Adolescent Health* 60(5). [Available here](#).

²⁵ Belong To (2021) *LGBTI+ Life in Lockdown: One Year Later*. Dublin: Belong To. [Available here](#).

misinformation have warned about the deliberate conflation of age-appropriate information relating to LGBTQ+ people and identities, and accusations of “grooming” and “sexualising” children.²⁶ As such, it is vital that content-rating processes, particularly in a case where it is determined algorithmically, do not automatically deem LGBTQ+-related content to be inappropriate for children and young people.

Question 12: Parental control features

What requirements should the Code have in relation to parental control features?

For a more detailed response to this question, we direct to the comprehensive submission compiled by the Children’s Rights Alliance.

Similarly to age verification, relating to parental controls, there are a number of additional considerations to be taken in the case of LGBTQ+ young people. As stated previously, international research shows that LGBTQ+ young people use social media to seek community, and to look for the safe spaces and information they may not have access to in real life.²⁷ In an Irish context, this source of community and support is particularly important for LGBTQ+ youth, 56% of whom live in home environments that are not supportive of their LGBTQ+ identity.²⁸

As a result, consideration of the above should be given to parental control measures which require the input and/or consent of a parent, carer or guardian for a young person to create a social media account, and/or access certain forms of content, balanced against rights enshrined under the UN Convention on the Rights of the Child to freedom of expression (article 13); freedom of thought, conscience and religion (article 14); freedom of association (article 15); and access to appropriate information (article 17).

²⁶ Fitzgerald, C. (2023) ‘Explainer: Why is the far-right targeting Ireland’s libraries?’, *The Journal.ie*. [Available here](#).

²⁷ Steinke, J. Root-Bowman, M. Estabrook, S. Levine, D. Kantor, L. (2017) ‘Meeting the Needs of Sexual and Gender Minority Youth: Formative Research on Potential Digital Health Interventions’, *Journal of Adolescent Health* 60(5). [Available here](#).

²⁸ Belong To (2021) *LGBTI+ Life in Lockdown: One Year Later*. Dublin: Belong To. [Available here](#).

Question 13: Media literacy measures and tools

What requirements should the Code contain to ensure that VSPS provide for effective media literacy measures and tools?

For a more detailed response to this question, we direct to the comprehensive submission compiled by the Children's Rights Alliance.

In 2022, the Reuters *Digital News Report* found that online sources have remained the number one source of news information among the Irish public, a position retained since 2015.²⁹ In 2022, 83% of the Irish public sourced news from online platforms including social media, compared to 63% accessing news from TV and 27% accessing news from print media. 51% of the Irish public sourced news from social media, with the leading platform being Facebook (33%), followed by WhatsApp (20%) and YouTube (20%).

As stated previously, the European Digital Media Observatory (EDMO), an international organisation that seeks to analyse disinformation, reported in May of this year that “mis- and disinformation targeting the LGBTQ+ community is one of the most present and consistent in the European Union”.³⁰ As such, the media literacy measures and tools should be designed in consultation with the LGBTQ+ sector, so as to ensure that they are robust and comprehensive in addressing disinformation relating to the LGBTQ+ community. The approach to designing these measures and tools should also be guided by research and best-practice in countering disinformation relating to the LGBTQ+ community, and other marginalised groups.

Question 14: Terms and conditions

How should we ask VSPS providers to address online harms in their terms and conditions in the Code, including the harms addressed under Article 28b?

Terms and conditions should be written in plain, accessible language that can be easily understood by children and young people. Youth consultation in developing these terms and conditions would be a meaningful consideration in achieving this goal.

²⁹ Reuters (2022) *Ireland: Digital News Report*. [Available here](#).

³⁰ Panizio, E. and Canetta, T. (2023) ‘Rights in the time of conspiracies and fake news: disinformation against LGBTQ+ in the EU’. European Digital Media Observatory: Italy. [Available here](#).

As detailed by the 5Rights Foundation, it is vital that terms and conditions:

- use simple language.
- aid comprehension.
- be concise.
- be presented in multiple formats for different age ranges.
- be prominent and easy to find.
- be presented at the right moments in a user journey.
- consider the diverse needs of young people.
- not assume adult involvement.
- cater for children with accessibility needs.
- ensure that consent must be obtained and sought, not assumed.
- ensure users are given meaningful choices.³¹

Question 15: Content moderation

How should we ask VSPS providers to address content moderation in the Code?

Effective content moderation ensures that the burden is not primarily placed on users to address harmful content through flagging mechanisms.³²

As outlined in relation to content rating, it is important that the Code require social media platforms follow best-practice guidelines in content detection and moderation, that are informed by LGBTQ+ identities and experiences. Over the past number of years, media outlets have reported that the VSPS, TikTok, has censored or suppressed LGBTQ+ content, creators and hashtags, despite this content not being in breach of community guidelines.³³ As such, it is vital that automated content detection and moderation processes do not automatically deem certain LGBTQ+-related terms or phrases to be in potential breach of community guidelines.

³¹ 5Rights Foundation, Tick to Agree Age appropriate presentation of published terms September 2021, 10-22. [Available here](#).

³² 5Rights Foundation, Tick to Agree Age appropriate presentation of published terms September 2021, 34. [Available here](#).

³³ Bacchi, U. (2020) 'TikTok apologises for censoring LGBT+ content'. Reuters. [Available here](#).

Question 16: Complaint-handling, resolution and reporting

What requirements should the Code include about procedures for complaint-handling and resolution, including out-of-court redress or alternative-dispute resolution processes?

In its 2021 General Comment on children’s rights in relation to the digital environment, the UN Committee on the Rights of the Child set out a number of recommendations relating to complaint handling and resolution.³⁴ It recommended that judicial and non-judicial remedial mechanisms be made available for children in relation to digital rights violations, and that these mechanisms be “widely known and readily available to all children”. Additionally, the Committee recommended that complaint handling be “swift”, and that these mechanisms be “free of charge, safe, confidential, responsive, child-friendly and available in accessible formats”.

How frequently should VSPS providers be obliged to report to the Commission on their complaint handling systems and what should those reports contain?

For the above conditions to be met, it is vital that the complaint handling mechanisms of VSPS providers are quick and effective, are to be addressed by the platform within a maximum time-period, are transparent for users, and are bound by annual reporting requirements to Coimisiún na Meán.

Question 18: Risk assessments and safety by design

What approach do you think the Code should take to risk assessments and safety by design?

The Online Safety Code should integrate key principles of the Council of Europe’s ‘Guidelines to Respect, Protect and Fulfil the Rights of the Child in the Digital Environment’.³⁵ Namely, these key principles include the requirement that safety by design, privacy by design, and privacy by default, taking into account the best interests of the child. Additionally, VSPS platforms should be required to regularly conduct child-

³⁴ UN Committee on the Rights of the Child, General Comment no 25 (2021) on children’s rights in relation to the digital environment, CRC/C/GC/25, para 44-46. [Available here.](#)

³⁵ Council of Europe, ‘Recommendation CM/Rec(2018)7 of the Committee of Ministers to Member States on Guidelines to respect, protect and fulfil the rights of the child in the digital environment’ (COE 2018) 10. [Available here.](#)

rights impact assessments, and bound by reporting requirements detailing mitigation measures required to address these risks, and the progress of these mitigation measures.

Question 23: Transition periods and timeframes

Should the Code have a transition period or transition periods for specific issues?

The Online Safety Code should come into force as soon as possible, without delay.

As outlined in earlier sections, anti-LGBTQ+ content is common on social media platforms. In 2022, 87% of LGBTQ+ youth reported seeing or experiencing anti-LGBTQ+ hate and harassment on social media in the past year.³⁶ The European Digital Media Observatory (EDMO) reported in May of this year that “mis- and disinformation targeting the LGBTQ+ community is one of the most present and consistent in the European Union”.³⁷ Research conducted in 2021 found that LGBTQ+ people experience 50% more online hate and harassment than any other minority group.³⁸

Belong To supports the recommendation of the Children’s Rights Alliance that the transition period should be as short as possible, taking the example of the UK Children’s Code, which provided for a one-year transition period.³⁹

³⁶ Pizmony-Levy, O. (2022) *The 2022 Irish School Climate Survey*. Research Report. Global Observatory of LGBTQ+ Education and Advocacy. Dublin and New York: Belong To and Teachers College, Columbia University. [Available here](#).

³⁷ Panizio, E. and Canetta, T. (2023) ‘Rights in the time of conspiracies and fake news: disinformation against LGBTQ+ in the EU’. European Digital Media Observatory: Italy. [Available here](#).

³⁸ ADL Centre for Technology & Society (2021) *Online Hate and Harassment: The American Experience*. ADL: New York. [Available here](#).

³⁹ Information Commissioner’s Office, ‘Age Appropriate Design: A Code of Practice for Online Services’ 21. [Available here](#).

BODYWHYS

The Eating Disorders Association of Ireland

Online Safety Code for

Video-Sharing Platform Services

Submission to Coimisiún na Meán

Contents

About Bodywhys	1
About eating disorders	1
Statistics	2
Role of social media	2
Current submission.....	2
Responses to Call for Input questions	3
Additional comments	7
Conclusion	7
Appendix 1: Bodywhys online harms survey feedback	7

About Bodywhys

Founded in 1995, Bodywhys – The Eating Disorders Association of Ireland - is the national voluntary organisation supporting people affected by eating disorders and their families. Bodywhys provides a range of non-judgemental listening, information and support services, professional training, literature, podcasts and webinars. Other aspects of the organisation's work include developing professional resources and collaborating with social media companies to respond to harmful online content and working with the mainstream media to create awareness about eating disorders. Bodywhys develops evidence-based programmes to promote positive body image and social media literacy in children and adolescents, as well as school talks and educational resources. Bodywhys is the support partner to the HSE's National Clinical Programme for Eating Disorders (NCP-ED), which delivers specialist public services in the Republic of Ireland.

About eating disorders

Eating disorders are serious and complex mental illnesses that pose risks to a person's physical, psychological, and emotional health and they lead to increased risk of suicide and mortality.¹ They often require medical intervention and ongoing treatment to help a person move towards recovery, with specialised care being key. Early assessment and evidence-based treatment improves the likelihood of recovery.² Specialist outpatient treatment represents the most effective and fastest way for most people with eating disorders to recover.³ Eating disorders involve behavioural, cognitive, emotional and physical aspects, which is why it takes time to recover and find treatment that works for each individual person's experience. Current diagnosable eating disorders include anorexia nervosa, bulimia nervosa, binge eating disorder, avoidant/restrictive food intake disorder (ARFID) and other specified feeding or eating disorder (OSFED). Eating disorders are not a lifestyle choice, a phase or a diet. Eating disorders affect a broad range of people from young people to adults, women and historically overlooked groups such as boys and men, members of the LGBTQIA+ community and those who are neurodivergent.

Statistics

- Based on epidemiological projections, the NCP-ED estimates that 188,895 people in Ireland will experience an eating disorder at some point in their lives.⁴ It is estimated that approximately 1,757 new cases occur in Ireland each year in the 10-49 age group
- Media reporting in 2022 identified an almost five-fold increase in cases of eating disorders at the Children's Hospital in Tallaght over the past eight years⁵
- In July 2023, the Health Research Board (HRB) reported that the number of child and adolescents admissions for eating disorders more than doubled in the last 5 years, from 33 in 2018 to 80 in 2022⁶.

Role of social media

Social media is a tool for communication and expression and it can be a space for body acceptance. Some people experiencing eating disorders use social media to connect with others genuinely and positively in a similar situation or to share aspects of their illness and recovery. Whilst this is sometimes helpful, there is a fine line between what's helpful and harmful. Aspects of social media can pose challenges to recovery. For example, underrepresentation of the diversity of bodies and ethnicities, misinformation and the promotion of incorrect or harmful recovery strategies, diet culture content, anti-recovery content that's easy to access, content that induces competition and comparisons in recovery, stigmatisation of people in larger bodies, progress and shaming of relapses, algorithms repeating the same trends, risky challenges and inaccurate information about mental health.^{7,8}

Current submission

Bodywhys welcomes the inclusion of online harmful content relating to eating disorders in recent legislation, the Online Safety and Media Regulation Act 2022. We also welcome the opportunity to input into the development of Ireland's first binding Online Safety Code for video-sharing platforms (VSPS). We have answered questions which are most relevant to our area of knowledge, work and experience.

Responses to Call for Input questions

Question 1

Ireland's first Online Safety Code should be a template for how online safety issues are addressed by Coimisiún na Meán and in turn, VSPS. This includes efforts to reduce harm, improve accountability, standards, transparency, actions and outcomes, develop long-term objectives, with a focus on ultimately acting in the best interests of those who access, view and interact with online content. The Code may wish take into consideration marginalised groups across race and ethnicity, sexual, gender and socioeconomic backgrounds and those who are differently abled or who experience chronic health issues.⁹

Bodywhys recommends that eating disorders are recognised as a form of online harms in the forthcoming Code. Online harmful content about eating disorders is typically described as 'pro-anorexia' or 'pro-ana'. This material tends to focus on endorsing or promoting specific eating disorder behaviours, such as risky food and weight behaviours. An extensive body of research evidence has identified key implications of exposure to pro-ana content, including how it affects someone's thoughts and feelings, weight and eating behaviours, the level of pressure and stigma they feel under and that the communities are not supportive. Several factors may underline the risks for users of pro-anorexia websites including, increasing availability and accessibility, the type of content and frequency of visits and the vulnerability of the user.¹⁰ Prior to social media, pro-ana material was typically found on message boards, websites or blogs. It must be addressed from a regulatory perspective because of its detrimental impact on users.

Question 2

For a variety of reasons, some children may not have the confidence and knowledge to express, to a trusted adult, that they've seen, received or experienced something harmful via the online space. A few of what might happen after speaking up is also a worry. It is imperative to design and enforce stringent measures to protect their emotional and psychological well-being. In its submission to the Call for Inputs, the Children's Rights Alliance (CRA) helpfully outline approaches considered by Children Online: Research and Evidence (CO:RE) and in Australia and in the United Kingdom.¹¹

Question 3

- [Urgent Responsibility to Reduce Harms Posed by Social Media on risk for Eating Disorders: An Open Letter to Facebook, Instagram, TikTok, and Other Global Social Media Corporations.](#)
- [Health advisory on social media use in adolescence \(apa.org\).](#)
- [Surgeon General Issues New Advisory About Effects Social Media Use Has on Youth Mental Health | HHS.gov](#)
- [Deadly by Design — Centre for Countering Digital Hate - TikTok pushes harmful content promoting eating disorders and self-harm into young users' feeds | CCDH \(counterhate.com\)](#)
- [Investigating Risks and Opportunities for Children in a Digital World \(lse.ac.uk\)](#)
- [The impact of digital experiences on adolescents with mental health vulnerabilities | Media@LSE](#)
- [New research from Butterfly Foundation highlights impact of social media - Butterfly Foundation](#)
- [Online advertising and eating disorders - Beat \(beateatingdisorders.org.uk\)](#)
- [Global Kids Online | Children's rights in the digital age](#)
- [Can the Metaverse Be Good for Youth Mental Health? Youth-Centred Strategies \(jedfoundation.org\)](#)
- [Insta pro-eating disorder bubble April '22 \(reset.tech\)](#)

- [Childrens-Commissioner-for-England-Life-in-Likes \(childrenscommissioner.gov.uk\)](https://www.childrenscommissioner.gov.uk)
- [Policies to protect children from the harmful impact of food marketing: WHO guideline](#)

Question 4.1

We echo the comments of the CRA who highlight the legal, regulatory and children-centred reasons for introducing option 1, a very detailed, perspective Code.

Question 4.3

We agree with the Coimisiún’s suggestion for the Code to mirror provisions of the Digital Services Act (DSA), where possible. We further agree on the need for the Code to identify metrics about the timing and accuracy of moderation actions and decisions that apply to types of content.

Question 4.4

We live in an increasingly visual and device-centric world, where there is often appearance-based content, messages and advertising. In the era of social media, anyone can share messages related to diet, weight, exercise, food and/or bodies without any requirement to reference relevant qualifications or without information from credible sources. This means that complex topics can be broken down into overly simplistic messages, which can be absorbed by those who are vulnerable, as well as being unhelpful to those who are unwell or trying to recover from an eating disorder. Such messages alone do not cause individual cases of eating disorders, however some may not be age appropriate and contribute to a confusing environment and unrealistic health, fitness and weight goals and norms. The recent Dove video “Cost of Beauty: A Dove Film” profoundly captures how things can escalate and subsequently deteriorate for a person.¹²

Videos can spread rapidly, generate significant viewership, commentary, critiques, responses, traction and interest. Recent videos developed by Webwise through its #SilentWitness campaign show how ordinary social or

peer group situations can be taken out of context and misrepresented online.^{13,14} It is also concerning, as suggested by mainstream media reporting, that some people may turn to social media for mental health advice or to learn if they have a particular illness.¹⁵ As one clinician noted, 'some of the take-home messages they have picked up are reckless and potentially dangerous.'¹⁶

Pro-anorexia video content has been identified on a range of VSPS.^{17,18,19,20,21} This is no less problematic than other manifestations of pro-ana material, such as text-based information, lyrics, blogs or extreme diets. Currently, feedback received by Bodywhys suggests that both the way some videos are experienced by, and suggested to, users can be notably problematic. Overall, this material is experienced as intrusive, hard to navigate, with limited success after requesting that it is actioned. For instance, 'What I eat in a day' videos which may focus on very restrictive food intake patterns or diets.

Question 7

We agree with the summary of the ten measures outlined in Article 28b.3 of the Audiovisual Media Services Directive (AVMSD) in the Coimisiún's Call for Inputs document.

Question 13

Creating and developing media literacy tools is not the responsibility of one group. As outlined by the United States Surgeon General and the Jed Foundation, it requires prioritisation and input from researchers, funders, policymakers, school and community organisations.^{22,23} Media literacy tools and resources available through VSPS must be age appropriate, easy to access and navigate.

Question 14-16, 18, 23

We agree with the responses outlined by CRA to these questions.

Additional comments

Some platforms may produce statistical reports which document the removal of harmful content by time and volume, such as within a community guidelines enforcement framework. Whilst useful overall, how this is reported and organised may sometimes be confusing. Previously, with TikTok for instance, 'dangerous acts', which were not defined as a mental health issue, were bundled together with suicide and self-harm as a combined category. It appears that, only since October 2022, has TikTok combined eating disorders with self-harm and suicide in its reporting and separated out dangerous acts. Where a removal policy or classification changes, VSPS must clarify the implications of this in how they share data about the removal of harmful online content.

We welcome the range of points noted by, and advocacy of, the Academy for Eating Disorders (AED), American Psychological Association (APA), the United States Surgeon General and the Jed Foundation whose work is noted in our submission and whose voices reflect the need for substantial and systemic change.

Conclusion

We believe that there is a noticeable, yet unsurprising gap between what VSPS stated they have done and what people with direct experience of eating disorders are encountering online. It is unknown at this point whether this is a consequence of inconsistent moderation, the algorithm or other factors. We look forward to the development of the Ireland's first Online Safety Code improving how people affected by eating disorders navigate online spaces.

Appendix 1: Bodywhys online harms survey feedback

We share the following sample of quotes from voices of lived experience, family members and friends who completed our online harms survey which ran from April-June 2023. We asked people to describe their experiences of encountering harmful online content related to eating disorders, and the effectiveness of the reporting process available from social media platforms.

“I found it harmful as I’ve seen videos promoting calorie restriction and tips on how to lose weight quickly and dangerously.”

“Specifically, on TikTok and Instagram, what I eat in a day videos promoting extremely restrictive low calorie diets. ‘Recovery’ accounts that promote extreme exercise, for example people claiming to be in eating disorder recovery but really have just turned from one ED to another, like from a restrictive ED to an obsessive exercise focused ED, anything from marathon training to gym obsessed. These accounts are harmful because they're suggesting that recovery should look this certain way, still controlling the way your body looks through a different means.”

“I did not seek it out. I tried to block that type of content but no luck. I feel like it is coming into my personal space and head space repeatedly.”

“It had a lot of tips and tricks encouraging other sufferers like myself to want to relapse instead of trying to recover.”

“Tips on how to avoid eating around family, content shaming fat people and encouraging anorexic behaviour and tips to make yourself sick after eating.”

End of submission

Jacinta Hastings,

Chief Executive Officer,

Bodywhys - The Eating Disorders Association of Ireland,

Postal Address PO Box 105, Blackrock, Co. Dublin.

Office Tel: [REDACTED]

ceo@bodywhys.ie

Think Bodywhys CLG, trading as Bodywhys - The Eating Disorders Association of Ireland, is a company limited by guarantee, registered in Ireland with a registered office at 18 Upper Mount Street, Dublin 2 and registered company number 236310.

Bodywhys is also a charity (Charity Reg. No. 20034054) and holds CHY number 11961.

Web: www.bodywhys.ie

Office Tel: 01 2834963

Helpline: 01 2107906

Email support: [REDACTED]

¹ Academy for Eating Disorders (2021) *AED Report, Eating Disorders: A Guide Medical Care, 4th Edition*.

² NICE (2017) *Eating disorders: Recognition and treatment*. National Institute for Health and Care Excellence

³ NICE (2017) *Ibid*

⁴ Health Service Executive (2018) *Eating Disorder Services: HSE Model of Care for Ireland*. Dublin: Health Service Executive

⁵ O’Keeffe, C. (2022) Four-fold jump in youth eating disorder cases, accelerated by pandemic. *Irish Examiner*. Accessed 25 August 2023. Available from: <https://www.irishexaminer.com/news/arid-40905803.html>

⁶ Daly, A., Lynn, E. (2023) Annual Report on the Activities of Irish Psychiatric Units and Hospitals, 2022. Dublin: Health Research Board. Accessed 25 August 2023. Available from: <https://www.hrb.ie/data-collections-evidence/psychiatric-admissions-and-discharges/publications/publication/activities-of-irish-psychiatric-units-and-hospitals-2022/returnPage/1/>

⁷ Au, E.S. & Cosh, S.M. (2022) Social media and eating disorder recovery: An exploration of Instagram recovery community users and their reasons for engagement. *Eating Behaviours*, 46:101651. doi: 10.1016/j.eatbeh.2022.101651.

⁸ Pruccoli, J., De Rosa, M., Chiasso, L., Perrone, A., Parmeggiani, A. (2022) The use of TikTok among children and adolescents with Eating Disorders: experience in a third-level public Italian center during the SARS-CoV-2 pandemic. *Italian Journal of Paediatrics*, 48(1), 138. doi: 10.1186/s13052-022-01308-4. PMID: 35907912; PMCID: PMC9338669.

-
- ⁹ American Psychological Association (2023) Health advisory on social media use in adolescence. Accessed 25 August 2023. Available from: <https://www.apa.org/topics/social-media-internet/health-advisory-adolescent-social-media-use>
- ¹⁰ Bond, E. (2012) *Virtually anorexic – Where’s the harm? A research study on the risks and pro-anorexia websites*. Accessed 31 August 2023. Available from: <https://www.thechildrensmediafoundation.org/wp-content/uploads/2014/02/Bond-2012-Research-on-pro-anorexia-websites.pdf>
- ¹¹ Children’s Rights Alliance (2023) Submission to Coimisiún na Meán on Developing Ireland’s First Binding Online Safety Code for Video-Sharing Platform Services.
- ¹² Dove (2023) Cost of Beauty: A Dove Film | Dove Self-Esteem Project [Video]. YouTube. Accessed on 16 August 2023. Available from: <https://www.youtube.com/watch?v=2ngESNoacxM>
- ¹³ Webwise (2023) #SilentWitness [Video]. YouTube. Accessed on 23 August 2023. Available from: <https://www.youtube.com/watch?v=Haf1CwB8MBY>
- ¹⁴ Webwise (2023) #SilentWitness: A Snapshot. [Video]. YouTube. Accessed on 23 August 2023. Available from: https://www.youtube.com/watch?v=d-5S_dLNUEM
- ¹⁵ Murphy Kelly, S. (2023) Teens are using social media to diagnose themselves with ADHD, autism and more. Parents are alarmed. *CNN*. Accessed 24 August 2023. Available from: <https://edition.cnn.com/2023/07/20/tech/tiktok-self-diagnosis-mental-health-wellness/index.html>
- ¹⁶ Noctor, C. (2023) Colman Noctor: Young people are using TikTok to self-diagnose mental health conditions. *Irish Examiner*. Accessed 24 August 2023. Available from: <https://www.irishexaminer.com/lifestyle/parenting/arid-41006648.html>
- ¹⁷ Syed-Abdul, Fernandez-Luque, L., Jian, W.S. et al. (2013) Misleading health-related information promoted through video-based social media: Anorexia on YouTube. *Journal of Medical Internet Research*, 15(2), e30.
- ¹⁸ Yom-Tov, E., Fernandez-Luque, L., Weber, I. & Crain, S.P. (2012) Pro-anorexia and pro-recovery photo sharing: a tale of two warring tribes. *Journal of Medical Internet Research*, 14(6), :e151.
- ¹⁹ De Choudhury, M. (2015) Anorexia on Tumblr: A characterization study. Proceedings of the 5th International Conference on Digital Health 2015.
- ²⁰ Chancellor, S., Jerry Lin, Z. & Goodman, E.L. et al. (2016) Quantifying and predicting mental illness severity in online pro-eating disorder communities.

Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing.

²¹ Arseniev-Koehler, A., Lee, H. & McCormick, T. (2016) #Proana: Pro-eating disorder socialization on Twitter. *Journal of Adolescent Health*, 58(6), 659-664.

²² US Department of Health and Human Services (2023) Surgeon General Issues New Advisory About Effects Social Media Use Has on Youth Mental Health.

Accessed 25 August 2023. Available from:

<https://www.hhs.gov/about/news/2023/05/23/surgeon-general-issues-new-advisory-about-effects-social-media-use-has-youth-mental-health.html>

²³ Jed Foundation (2023) Can the Metaverse be good for youth mental health?

Accessed 25 August 2023. Available from: <https://jedfoundation.org/metaverse-and-youth-mental-health/>



Online Safety Code

Response to Coimisiún na Meán Online
Safety Code Call for Inputs

04 September 2023



Coimisiún um
Iomaíocht agus
Cosaint Tomhaltóirí

Competition and
Consumer Protection
Commission

Introduction

The Competition and Consumer Protection Commission ('the CCPC') welcomes the opportunity to respond to the Coimisiún na Meán ('CnaM') call for input into the development of Ireland's first binding Online Safety Code ('the Code').

The CCPC has a statutory function under Section 10(3)(a) of the Competition and Consumer Protection Act 2014 to provide advice to policymakers on matters likely to impact on consumer protection and welfare, or competition and the CCPC's submission reflects this mandate.

We are supportive of the decision by CnaM to focus the first online safety code on video-sharing platform services (VSPS) providers and to make sure VSPS providers take measures to address online harms more effectively.

Our response to this call for input will focus on question 8 from section 5.1 (Online Safety Features for Users) and question 19 from section 5.3.4 (Cooperation with other Regulators, Bodies) of the call for input paper.

Content containing commercial communication

It is important that users of VSPS are made aware when they are being presented with commercial communication. The eCommerce Directive and the Unfair Commercial Practices Directive contain provisions which are intended to ensure that consumers are informed when they are being presented with commercial communications and to protect them against misleading advertising or marketing with the potential to create consumer detriment. Therefore, we welcome the intention of the CnaM for the code to require VSPS providers to implement a feature for content creators to declare when the videos they upload contain commercial communication.

As noted in the call for input document, last year the CCPC published the results of research we conducted on online consumer behaviour and influencer marketing¹. The research found that consumers may be over confident in their ability to recognise when posts by influencers are in fact marketing and they may be more vulnerable to misleading marketing than they think. Nearly 24% of consumers who responded to a survey as part of the research stated that they felt misled about a product they had purchased as a result of an influencer promoting it online. This equates to 4.6% of the adult population.

A key finding from the research was that a significant portion of the posts with commercial content that we analysed were either not labelled at all or not sufficiently labelled. When we engaged directly with consumers and influencers we found that there was widespread agreement amongst both groups that clear guidance would be beneficial for everyone. In light of our findings, we concluded that the most appropriate approach to regulating influencer marketing in the Irish context is hybrid in nature encompassing: strengthened guidance; education of consumers, influencers, brands and agents; increased responsibility for platforms, and compliance and enforcement. The proposal to introduce a feature that allows users to declare when videos contain advertising or other types of commercial communication, as set out in the call for input, is aligned with the approach to regulating influencer marketing as recommended in the CCPC report.

As part of our research we reviewed international approaches to regulating influencers and identified practices used elsewhere that may be of relevance to CnaM in developing its Code. For example, in Denmark research has found that where hashtags or the “Paid Partnership” tag is highlighted, rather than just presented in standard text, it is more effective in allowing the consumer to correctly identify commercial content.

Cooperation between regulators

The CCPC notes the references in section 5.3.4 to cooperation by CnaM with other bodies in Ireland, including the CCPC. The CCPC further notes that the call for inputs poses the question of how regulators and agencies can cooperate to implement the Code.

¹ Available here: [CCPC influencer marketing research - CCPC Business](#)



The CCPC would be happy to engage further with CnaM on the influencer aspects of the Code as well as the broader issue of interaction between our two agencies.

We hope the points that we have raised are of assistance to CnaM and we look forward to engaging further on the development of the Code.

ENDS

Submission to Coimisiún na Meán on Developing Ireland's First Binding Online Safety Code for Video-Sharing Platform Services

September 2023



Founded in 1995, the Children’s Rights Alliance unites over 140 members working together to make Ireland one of the best places in the world to be a child. We change the lives of all children in Ireland by making sure that their rights are respected and protected in our laws, policies and services.

Accompaniment Support Service for Children (A.S.S.C.)
Alcohol Action Ireland
Amnesty International Ireland
An Cosán
Anew
Anne Sullivan Foundation
Aoibhneas
Archways
AsIAm
Association of Occupational Therapists of Ireland (AOTI)
Association of Secondary Teachers Ireland (ASTI)
ATD Fourth World – Ireland Ltd
Atheist Ireland
Barnardos
Barretstown Camp
Bedford Row Family Project
BeLonG To Youth Services
Bodywhys
Catholic Guides of Ireland
Child Law Project
Childhood Development Initiative
Children in Hospital Ireland
Children’s Books Ireland
Children’s Grief Centre
Clarecare
COPE Galway
Cork Life Centre
Cork Migrant Centre
Crann Centre
Crosscare
CyberSafeKids
Cycle Against Suicide
Dalkey School Project National School
Daughters of Charity Child and Family Service
Dental Health Foundation of Ireland
Department of Occupational Science and Occupational Therapy, UCC
Disability Federation of Ireland
Doras
Down Syndrome Ireland
Dublin Rape Crisis Centre
Dyslexia Association of Ireland
Dyspraxia/DCD Ireland
Early Childhood Ireland
Early Learning Initiative (National College of Ireland)
Educate Together
EPIC
Equality for Children
Extern Ireland
FamiliBase
Féach
Focus Ireland
Foróige
Gaeloideachas
Galway Traveller Movement
Good Shepherd Cork
Grow It Yourself
Helium Arts
Immigrant Council of Ireland
Inclusion Ireland
Institute of Guidance Counsellors
Irish Aftercare Network
Irish Association for Infant Mental Health
Irish Association of Social Workers
Irish Congress of Trade Unions (ICTU)
Irish Council for Civil Liberties (ICCL)
Irish Foster Care Association
Irish Girl Guides
Irish Heart Foundation
Irish National Teachers Organisation (INTO)
Irish Penal Reform Trust
Irish Primary Principals’ Network
Irish Refugee Council
Irish Second Level Students’ Union (ISSU)
Irish Society for the Prevention of Cruelty to Children
Irish Traveller Movement
Irish Youth Foundation
iScoil
Jigsaw
Katharine Howard Foundation
Kerry Diocesan Youth Service (KDYS)
Kids’ Own Publishing Partnership
Kinship Care
Leap Ireland
Let’s Grow Together! Infant and Childhood Partnerships CLG.
LGBT Ireland
Meath Women’s Refuge & Support Services
Mecpaths
Mental Health Reform
Mercy Law Resource Centre
Migrant Rights Centre Ireland
Mothers’ Union
Museum of Childhood Ireland
Music Generation
My Mind
My Project Minding You
National Childhood Network
National Council for the Blind of Ireland
National Forum of Family Resource Centres
National Parents Council Post Primary
National Parents Council Primary
National Youth Council of Ireland
New Directions
Novas
One Family
One in Four
Parents Plus
Pavee Point
Peter McVerry Trust
Prevention and Early Intervention Network
Psychological Society of Ireland
Rainbow Club Cork
Rainbows Ireland
Rape Crisis Network Ireland (RCNI)
Realt Beag/Ballyfermot Star
Respond Housing
SAFE Ireland
Saoirse Housing Association
SAOL Beag Children’s Centre
School of Education UCD
Scouting Ireland
Sexual Violence Centre Cork
Simon Communities of Ireland
SIPTU
Social Care Ireland
Society of St. Vincent de Paul
SPHE Network
SpunOut.ie
St. Nicholas Montessori College
St. Nicholas Montessori Teachers’ Association
St. Patrick’s Mental Health Services
TASC
Teachers’ Union of Ireland
Terenure College Rugby Football Club
The Ark, A Cultural Centre for Children
The Irish Red Cross
The Jack and Jill Children’s Foundation
The UNESCO Child and Family Research Centre, NUI Galway
The Wheel
Transgender Equality Network Ireland
Traveller Visibility Group Ltd
Treoir
UNICEF Ireland
Variety – the Children’s Charity of Ireland
Women’s Aid
Young Ballymun
Young Social Innovators
Youth Advocate Programme Ireland (YAP)
Youth Work Ireland

Children’s Rights Alliance

7 Red Cow Lane, Smithfield, Dublin 7, Ireland

Ph: +353 1 662 9400

Email: info@childrensrighths.ie

www.childrensrighths.ie

© 2023 Children’s Rights Alliance – Republic of Ireland Limited

The Children’s Rights Alliance is a registered charity – CHY No. 11541

Introduction

The Children’s Rights Alliance unites over 140 organisations working together to make Ireland one of the best places in the world to be a child. We change the lives of all children in Ireland by making sure that their rights are respected and protected in our laws, policies and services. We identify problems for children. We develop solutions. We educate and provide information and legal advice on children's rights.

The Children’s Rights Alliance is also a member and National Partner of Eurochild, the largest network of organisations and individuals working with and for children in Europe. Eurochild works closely with the European Union, as protecting children’s rights is among the EU’s aims and values.

The Children’s Rights Alliance welcomes the opportunity to make a written submission to Coimisiún na Meán on developing Ireland’s first binding Online Safety Code for video-sharing platform services.

Children make up one third of global online users.¹ Results from a National Survey of Children, their Parents and Adults regarding Online Safety conducted between December 2019 and October 2020, found that 62 per cent of children and young people in Ireland, aged nine to 17 years use social media.² This rises to 90 per cent of 15 to 17 year olds.³ While the online world brings unparalleled opportunity to children to learn, create, connect and socialise, it also brings risk, including the loss of personal data, exposure to harmful content, cyberbullying, negative impacts on health and well-being, online grooming and extortion. In 2021, CyberSafeKids reported that a quarter of all children they worked with⁴ surveyed have seen or experienced something online in the last year that bothered them, with almost one third of those children having kept it to themselves rather than report it to their parents or someone else.⁵

The UN Committee on the Rights of the Child have acknowledged the increasing importance of the digital environment in that it ‘affords new opportunities for the realization of children’s rights, but also poses the risks of their violation or abuse.’⁶

While undoubtedly, the internet has significant positive impacts both for children and wider society, for too long legislation and policy have not kept pace with the evolution of the online world. This has left children and young people at risk and unprepared to appropriately navigate online platforms. The introduction of the Online Safety and Media Regulation Act 2022 and the Digital Services Act will pave the way for a new era of online regulation. Central to this is the introduction of the Online Safety Codes. We welcome to opportunity to take part in the consultation process and look forward to continued engagement to make the online world safer for children and young people.

¹ Unicef, *Children in the Digital World* (UNICEF 2017).

² National Advisory Council for Online Safety, *Report of a National Survey of Children, their Parents and Adults regarding Online Safety 2021* (2021) 8.

³ *ibid.*

⁴ CybersafeKids gathered data from 4,714 children over the 2021/22 academic year CyberSafeKids, *Annual Report 2021* (2022) 5.

⁵ CyberSafeKids, *Annual Report 2021* (2022) 3.

⁶ UN Committee on the Rights of the Child, General Comment no 25 (2021) on children’s rights in relation to the digital environment, *CRC/C/GC/25*, para 3.

Question 1: What do you think our main priorities and objectives should be in the first binding Online Safety Code for VSPS? What are the main online harms you would like to see it address and why?

Main priorities and objectives

The UN Committee on the Rights of the Child are clear that ‘the rights of every child must be respected, protected and fulfilled in the digital environment.’⁷ This should be one of the main objectives of the first binding Online Safety Code.

The UN Committee on the Rights of the Child provides that States should ‘take all appropriate legislative, administrative, social and educational measures to protect the child from all forms of physical or mental violence, injury or abuse, neglect or negligent treatment, maltreatment or exploitation, including sexual abuse...’⁸ Further it requires States to ‘ensure that relevant legislation provides adequate protection of children in relation to media and ICT’.⁹ This should be one of the main priorities and objectives of the Code.

The Committee has also recommend that ‘in all actions regarding the provision, regulation, design, management and use of the digital environment, the best interests of every child is a primary consideration.’¹⁰ The Council of Europe (COE) *Guidelines to Respect, Protect and Fulfil the Rights of the Child in the Digital Environment* provide that ‘in all actions concerning children in the digital environment, the best interests of the child shall be a primary consideration’ and further recommend that States should strike a balance between the child’s right to protection and their other rights to freedom of expression, participation and access to information.¹¹ The COE also acknowledges the differing levels of maturity and understanding of children at different ages and recommends that States recognise the evolving capacities of children which can mean that the ‘policies adopted to fulfil the rights of adolescents may differ significantly from those adopted for younger children’.¹²

The Committee on the Rights of the Child in 2013 issued a General Comment¹³ which has clarified the meaning of this principle in 2013 and stated that it has a three-fold meaning. The best interests principle is:

(a) A substantive right: The right of the child to have his or her best interests assessed and taken as a primary consideration when different interests are being considered in order to reach a decision on the issue at stake, and the guarantee that this right will be implemented whenever a decision is to be made concerning a child, a group of identified or unidentified children or

⁷ *ibid* para 4.

⁸ UN Convention on the Rights of the Child, A/ RES/44/25 (20 November 1989) Arts 19.1

⁹ UN Committee on the Rights of the Child, General Comment no 13(2011) on the right of the child to freedom from all forms of violence, CRC/C/GC/13 para41(g).

¹⁰ UN Committee on the Rights of the Child, General Comment no 25 (2021) on children’s rights in relation to the digital environment, CRC/C/GC/25 para 12.

¹¹ *ibid*, 12.

¹² Council of Europe, ‘*Recommendation CM/Rec(2018)7 of the Committee of Ministers to Member States on Guidelines to respect, protect and fulfil the rights of the child in the digital environment*’ (COE 2018) <<https://bit.ly/2Xp9hpE>> accessed 26 February 2021, 12.

¹³ UN Committee on the Rights of the Child (2013) *General Comment No. 14: The right of the child to have his or her best interests taken as a primary consideration (art. 3, para. 1)*, CRC/C/GC/14.

children in general. Article 3, paragraph 1, creates an intrinsic obligation for States, is directly applicable (self-executing) and can be invoked before a court.

(b) A fundamental, interpretative legal principle: If a legal provision is open to more than one interpretation, the interpretation which most effectively serves the child's best interests should be chosen. The rights enshrined in the Convention and its Optional Protocols provide the framework for interpretation.

(c) A rule of procedure: Whenever a decision is to be made that will affect a specific child, an identified group of children or children in general, the decision-making process must include an evaluation of the possible impact (positive or negative) of the decision on the child or children concerned. Assessing and determining the best interests of the child require procedural guarantees. Furthermore, the justification of a decision must show that the right has been explicitly taken into account. In this regard, States parties shall explain how the right has been respected in the decision, that is, what has been considered to be in the child's best interests; what criteria it is based on; and how the child's interests have been weighed against other considerations, be they broad issues of policy or individual cases.

The Alliance believes that one of the main priorities should be ensuring that the best interests of the child is a primary consideration in the Code. Alongside this the Code needs to acknowledge the 'evolving capacities of the child as an enabling principle that addresses the process of their gradual acquisition of competencies, understanding and agency' as 'risks and opportunities associated with children's engagement in the digital environment change depending on their age and stage of development.'¹⁴

Many of the digital services children and young people use are not designed to protect their rights or meet their needs.¹⁵ Research from the 5Rights Foundation found that 'pathways designed into digital services and products are putting children at risk' with designers tasked with 'optimising products and services for three primary purposes, all geared towards revenue generation.'¹⁶ The Online Safety Code presents a huge opportunity to embed the principle of safety by design into the Irish regulatory framework. It is important that this principle is not incorporated only to services specifically targeted to children and young people but to all the digital services children and young people are likely to actually access.¹⁷

Recommendations

- Ensure that the protection and fulfilment of children's rights online is a primary objective of the Code. In particular ensure that:
 - the right of the child to protection from abuse and exploitation online is embedded as a key principle.

¹⁴ UN Committee on the Rights of the Child, General Comment no 25 (2021) on children's rights in relation to the digital environment, CRC/C/GC/25, para 19

¹⁵ 5Rights Foundation, 'Design of Service' <<https://5rightsfoundation.com/our-work/design-of-service/>> accessed 4 September 2023.

¹⁶ 5Rights Foundation, September 2021 Pathways: A Summary Key findings and recommendations from Pathways: How digital design puts Children at Risk (2021) 7.

¹⁷ 5Rights Foundation, 'Design of Service' <<https://5rightsfoundation.com/our-work/design-of-service/>> accessed 4 September 2023.

- the best interests of every child is a primary consideration in all actions affecting them.
- Embed safety by design into the Online Safety Code as one of the main priorities and objectives.

Main Harms

The harms listed in the ‘Call for Inputs’ need to all form part of the online harms the code seeks to address.

The four areas set out in Article 28b of the Audio- Visual Media Services Regulation need to be addressed by the Online Safety Code:

1. Content that might impair the physical, mental or moral development of minors. This includes content that may be inappropriate for children, such as pornography.
2. Content that incites violence or hatred against a group of persons or a member of a group based on any of the grounds referred to in Article 21 of the European Charter of Fundamental Rights. These grounds include sex, race, colour, ethnic or social origin, genetic features, language, religion or belief, political or any other opinion, membership of a national minority, property, birth, disability, age or sexual orientation.
3. Content the dissemination of which constitutes a criminal offence under EU law.
4. Certain commercial communications that would not be permitted on broadcast or video-on-demand services. Commercial communications include advertising, sponsorship and product placement.

The categories of harm set out in the Broadcasting Act 2009 Act as amended by the Online Safety and Media Regulation Act 2022 also should be addressed by the online safety code:

- a. Harmful online content relating to 42 criminal offences under Irish law listed in Schedule 3 of the 2009 Act as amended
- b. Online content by which a person bullies or humiliates another person;
- c. Online content by which a person promotes or encourages behaviour that characterises a feeding or eating disorder;
- d. Online content by which a person promotes or encourages self-harm or suicide;
- e. Online content by which a person makes available knowledge of methods of self-harm or suicide;

Consultations with children and young people have shown that they are most disturbed by violent content online, it is key that the Code addresses this.¹⁸ Consideration should also be given to addressing Harmful Commercial Communications, particularly marketing of high fat, sugar and salt foods and breastmilk substitutes as there are heightened risks of, and harms associated with, commercial exploitation and negative impact on development and health that can occur as a result of marketing practices of these foods.¹⁹ Alcohol is one of the most heavily marketed products and as

¹⁸ EU Kids Online ‘EU Kids Online 2020: Survey results from 19 countries’ < <https://www.lse.ac.uk/media-and-communications/research/research-projects/eu-kids-online/eu-kids-online-2020>> accessed 4 September 2023, 142,149,151.

¹⁹ Irish Heart Foundation, Submission to Coimisiún na Meán Call For Inputs on Developing First Online Safety Code 2023.

such it is important that the new Online Safety Code protect children and the general public from the harmful commercial practices.²⁰

Online racism and hate speech against minority groups such as Traveller and Roma, Lesbian, Gay, Bisexual and Transgender (LGBT) young people, black and ethnic minorities people should form part of the harms that the Online Safety Code seeks to address. The new Online Safety code should seek to align with the Criminal Justice (Incitement to Violence or Hatred and Hate Offences) Bill 2022 that is currently going through the Oireachtas. This legislation will contain a list of protected characteristics such as race, colour, nationality, religion, national or ethnic origin, descent, gender, sex characteristics, sexual orientation, disability, and also specifically recognises Travellers as an ethnic group. It states – “references to “national or ethnic origin” include references to membership of the Traveller community (within the meaning of section 2(1) of the Equal Status Act 2000),”

Recommendations

- The harms listed in the ‘Call for Inputs’ need to all form part of the online harms the code seeks to address.
- Ensure that the Code addresses children’s exposure to violent content online.
- Consideration should also be given to addressing Harmful Commercial Communications, particularly marketing of high fat, sugar and salt foods and breastmilk substitutes and alcohol.

²⁰ Alcohol Action Ireland, Submission to Coimisiún na Meán Call For Inputs on Developing First Online Safety Code 2023.

Question 2: What types of online harms do you think should attract the most stringent risk mitigation measures by VSPs? How could we evaluate the impact of different types of harms e.g. severity, speed at which harm may be caused? Is there a way of classifying harmful content that you consider it would be useful for us to use?

Types of Harm

There needs to be stringent risk measures imposed on illegal and harmful content for children and young people. In particular there is a need to guarantee that the Code ensures that illegal material such as child sexual abuse materials, intimate images and material that incites hatred can be robustly and swiftly removed. Consultations with children and young people have shown that they are most disturbed by violent content online, it is key that the Code addresses this.²¹

The UN Convention on the Rights of the Child (UNCRC) guarantees all children the right to be protected from abuse, neglect and sexual exploitation.²² There is growing concern that online grooming, as well as the sharing of child exploitation material, increased online during Covid-19.²³ Irish teenagers are the fourth highest users in the EU for sexting.²⁴ A recent Report from the Children's Commissioner for England found that pornography consumption is widespread among children with 13 years old being the average age of first exposure.²⁵ A significant minority of children are first exposed to pornography at a very young age, 10 per cent of the over 1,000 young people surveyed had seen it by age nine, 27 per cent had seen it by age 11 and 50 per cent had seen it by 13.²⁶ The Children's Commissioner Report also found that children 'often stumble accidentally across pornography online'²⁷ and Twitter is the platform where the greatest number of children had seen pornography.²⁸ The majority, 79 per cent of 18-21 year olds surveyed had seen content involving sexual violence before turning 18, and 47 per cent of all those surveyed stated that 'girls expect sex to involve physical aggression such as airway restriction or slapping.'²⁹ The Children's Commissioner is conscious in the report that age verification will not be a 'silver bullet' for regulating online pornography as some teenagers, particularly older teenagers may continue to access online pornography.³⁰ The Code needs to take measures to address children's access to pornography and the advertising of prostitution.

The production and distribution of child sexual abuse and exploitation materials – whether in print, online, or livestreamed – represent a fundamental violation of children's rights and a breach of the UNCRC.³¹ These images effectively represent a digital crime scene, and people accessing these images directly contribute to the exploitation of child victims by creating demand and perpetuating the child's trauma. This abuse is ongoing until the image is removed. In 2021, the Internet Watch

²¹ EU Kids Online 'EU Kids Online 2020: Survey results from 19 countries' <<https://www.lse.ac.uk/media-and-communications/research/research-projects/eu-kids-online/eu-kids-online-2020>> accessed 4 September 2023, 142,149,151.

²² UN Convention on the Rights of the Child, A/ RES/44/25 (20 November 1989) Arts 19 and 34.

²³ Interpol, 'Child Sexual Exploitation and Abuse threats and trends: COVID-19 Impact' <<https://bit.ly/34unFDS>> accessed 1 February 2022.

²⁴ Dublin City University, 'Irish Teens the Fourth Highest in the EU for Sexting' <<https://bit.ly/3qTC2HK>> accessed 6 January 2022. See also: Raymond Arthur, 'Policing Youth Sexting in Ireland' (2019) 22(3) Irish Journal of Family Law 66.

²⁵ Children's Commissioner for England, 'A lot of it is actually just abuse' Young people and pornography' January 2023, 6-8.

²⁶ *ibid.*

²⁷ *ibid.*

²⁸ *ibid.*

²⁹ *ibid.*

³⁰ *ibid.*

³¹ UN Convention on the Rights of the Child, A/ RES/44/25 (20 November 1989) Arts 19 and 34.

Foundation received a 64 per cent increase in reports, of which 252,194 reports were confirmed as containing child sexual abuse and exploitation material.³² Similarly, Hotline.ie, the Irish national reporting centre for illegal online content, has experienced a dramatic increase in demand for its services – in 2021 it saw 29,794 reports, which was as many as the previous three years combined.³³ Despite this, Irish cases involving the distribution of child abuse material are taking up to 10 years for the State to complete, with the problem accelerating in recent years, as technological and data issues impede prosecutions.³⁴ There is a need to ensure that the Code provide for mechanisms for material to be removed swiftly.

Classifying Harmful Content

The Council of Europe has recommended that ‘states should co-operate with a view to promoting standardisation of content classification and advisory labels among countries and across stakeholder groups to define what is appropriate and what is inappropriate for children.’³⁵ There are a number of frameworks that could be considered.

CO:RE 4Cs classification

A key tool to identify risk and classification of harm is the 4Cs framework. The CO:RE 4Cs classification recognises that online risks arise when a child:

- Engages with and/or is exposed to potentially harmful content
- Experiences and/or is targeted by potentially harmful contact
- Witnesses, participates in and/or is a victim of potentially harmful conduct
- Is party to and/or exploited by a potentially harmful contract³⁶

The 4Cs classification ‘distinguishes between aggressive, sexual and value risks’ along with recognising important cross-cutting risks such as children’s right to privacy and fair treatment.³⁷

³² The Internet Watch Foundation, ‘IWF Annual Report 2021 – Face the Facts’ < <https://www.iwf.org.uk/about-us/who-we-are/annual-report-2021/> > accessed 22 November 2022.

³³ Hotline.ie, *2021 Annual Report (2022)* 7-8.

³⁴ Conor Gallagher, ‘Backlogs a dangerous flaw in child porn and abuse inquiries Resource and Data leave perpetrator at large – and child vulnerable to further abuse’ *The Irish Times*, 7 January 2020.

³⁵ Council of Europe, Guidelines to respect, protect and fulfil the rights of the child in the digital environment (2018) Recommendation CM/Rec(2018)7 of the Committee of Ministers, 29, para 121.

³⁶ CORE, ‘4 Cs of online risk: Short report & blog on updating the typology of online risks to include content, contact, conduct, contract risk’ <<https://core-evidence.eu/posts/4-cs-of-online-risk>> accessed 28 August 2023.

³⁷ *ibid.*

 CORE	Content Child as recipient	Contact Child as participant	Conduct Child as actor	Contract Child as consumer
Aggressive	Violent, gory, graphic, racist, hateful and extremist content	Harassment, stalking, hateful behaviour, unwanted surveillance	Bullying, hateful or hostile peer activity e.g. trolling, exclusion, shaming	Identity theft, fraud, phishing, scams, gambling, blackmail, security risks
Sexual	Pornography (legal and illegal), sexualization of culture, body image norms	Sexual harassment, sexual grooming, generation and sharing of child sexual abuse material	Sexual harassment, non-consensual sexual messages, sexual pressures	Sextortion, trafficking for purposes of sexual exploitation, streaming child sexual abuse
Values	Age-inappropriate user-generated or marketing content, mis/disinformation	Ideological persuasion, radicalization and extremist recruitment	Potentially harmful user communities e.g. self-harm, anti-vaccine, peer pressures	Information filtering, profiling bias, polarisation, persuasive design
Cross-cutting	Privacy and data protection abuses, physical and mental health risks, forms of discrimination			

38

Australian Classification Scheme

A classification scheme is in place in Australia where the Australian Online Safety Act (2021) defines content as either ‘class 1 material’ or ‘class 2 material’.³⁹ Class 1 material and class 2 material are defined by reference to Australia’s National Classification Scheme, which is also used for classification of films, computer games and other publications.

Class 1 material includes material that:

- ‘depicts, expresses or otherwise deals with matters of sex, drug misuse or addiction, crime, cruelty, violence or revolting or abhorrent phenomena in such a way that they offend against the standards of morality, decency and propriety generally accepted by reasonable adults to the extent that they should not be classified
- describes or depicts in a way that is likely to cause offence to a reasonable adult, a person who is, or appears to be, a child under 18 (whether the person is engaged in sexual activity or not), or
- promotes, incites or instructs in matters of crime or violence.’

Class 2 material is material that is, or would likely be, classified as either:

- ‘X18+ (or, in the case of publications, category 2 restricted), or
- R18+ (or, in the case of publications, category 1 restricted) under the National Classification Scheme, because it is considered inappropriate for general public access and/or for children and young people under 18 years old.’

The eSafety Commissioner works with online service providers to ensure access to Class 2 material, which is considered unsuitable for children and young people under 18, is restricted.

³⁸ CORE, ‘4 Cs of online risk: Short report & blog on updating the typology of online risks to include content, contact, conduct, contract risk’ <<https://core-evidence.eu/posts/4-cs-of-online-risk>> accessed 28 August 2023.

³⁹ Online Safety Act 2021 s106 and s107.

UK Classification Scheme

The UK Online Safety Bill currently going through the houses of parliament classifies online content that is harmful to children into two distinct categories – primary priority content, and priority content.⁴⁰ Primary priority content that is harmful to children includes pornographic content, content which encourages, promotes or provides instructions for an act of deliberate self-injury, suicide, or eating disorders or behaviour associated with eating disorders.⁴¹ Content also falls within the primary priority category if it consists only of text or text accompanied by identifying content which consist only of text, other identifying content which is not itself pornographic, a GIF which is not itself pornographic, an emoji or other symbol, or any combination of these.⁴² Priority content that is harmful to children includes content which is abusive and targets race, religion, sex, sexual orientation, disability, or gender reassignment, content which incites hatred against people of any of the above named groups, bullying content, content depicting real or realistic serious violence against a person (not limited to a real person) or animal, or serious injury of a person (not limited to a real person) or animal in graphic detail, as well as serious violence or injury against a fictional creature in graphic detail.⁴³ Also falling within the priority content category is content which encourages, promotes or provides instructions for a challenge or stunt highly likely to result in serious injury to the person who does it or someone else, content encouraging a person to self-administer a physically harmful substance or a substance in a quantity that is physically harmful.⁴⁴

Under section 63 of the UK Online Safety Bill the Office of Communications (OFCOM) must review the incidence on regulated user-to-user services, search services and combined services of content that is harmful to children, and the severity of harm that children in the UK suffer or may suffer as a result of such content.⁴⁵ Arising from this OFCOM must publish a report every three years at least on the outcome of the review and make recommendations as to whether changes are required for sections 61 and 62 covering primary priority content and priority content.⁴⁶

Recommendations

- Provide for mechanisms in the Code for child sexual abuse material to be removed swiftly.
- Take measures in the Code to address children’s access to pornography and the advertising of prostitution.
- Ensure that the Code provides that illegal material such as child sexual abuse materials, intimate images and material that incites hatred can be robustly and swiftly removed.
- Ensure that the Code addresses violent content online.

⁴⁰ UK Online Safety Bill, section 60(1).

⁴¹ *ibid* section 61(2) – (5).

⁴² *ibid* section 61(6).

⁴³ *ibid* section 62(1) – (7).

⁴⁴ *ibid* section 62(8) – (10).

⁴⁵ *ibid* section 63.

⁴⁶ *ibid*.

Question 3: Do you have reports, academic studies or other relevant independent research that would support your views? If you do, please share them with us with links to relevant reports, studies or research.

- 5Rights Foundation, But how do they know it is a child? Age Assurance in the Digital World
- 5Rights Foundation, Making Child Online Safety a Reality: Global Toolkit
- 5Rights Foundation, Pathways: A Summary Key findings and recommendations from Pathways: How digital design puts children at risk
- 5Rights Foundation, Tick to Agree Age appropriate presentation of published terms September 2021
- Child Rights Impact Assessment A tool to realise children's rights in the digital environment March 2021
- Council of Europe Handbook for policy makers on the rights of the child in the digital environment
- Council of Europe published its Recommendation, *Guidelines to Respect, Protect and Fulfil the Rights of the Child in the Digital Environment*
- Livingstone, Sonia (2016) A framework for researching Global Kids Online: understanding children's well-being and rights in the digital age. Global Kids Online. London School of Economics and Political Science, London, UK.
- Livingstone, Sonia and Pothong, Kruakae (2023) Child rights by design: guidance for innovators of digital products and services used by children. . Digital Futures Commission, 5Rights Foundation, London, UK.
- Livingstone, Sonia and Third, Amanda (2017) Children and young people's rights in the digital age: an emerging agenda. New Media & Society.
- Mukherjee, Sudeshna, Pothong, Kruakae and Livingstone, Sonia (2021) Child rights impact assessment: a tool to realise children's rights in the digital environment, Digital Futures Commission, 5Rights Foundation, London, UK.
- UN Committee on the Rights of the Child, General Comment no 25 (2021) on children's rights in relation to the digital environment, CRC/C/GC/25
- World Health Organisation, Policies to protect children from the harmful impact of food marketing: WHO guidelines (WHO 2013)
- Beating Eating Disorders UK, [Online advertising and eating disorders](#)
- Sonia Livingstone and Mariya Stoilova, ['The impact of digital experiences on adolescents with mental health vulnerabilities'](#) (2021)

Question 4: What approach do you think we should take to the level of detail in the Code? What role could non-binding guidance play in supplementing the Code?

The Online Safety Code should take the form of a detailed prescriptive Code. As noted in the ‘Call for Submissions’ this would allow the Code to ‘specify details in the measures we expect VSPS providers to take to address online harms.’

The UN Committee on the Rights of the Child in its General Comment no.25 on children’s rights in relation to the digital environment state that States should require the business sector to undertake children’s rights due diligence and child rights impact assessments and disclose them to the public with consideration of the ‘severe impacts of the digital environment on children.’⁴⁷ The UN Committee also state that States should require all businesses that affect children’s rights in relation to the digital environment to implement regulatory codes and frameworks to adhere to the highest levels of privacy and safety standards.⁴⁸ They also recommend that States encourage them to take accountability and measures to innovate in the best interests of the child.⁴⁹

A comprehensive legal and regulatory framework in this space should encompass both protective and preventive measures, prohibiting all forms of violence, exploitation and abuse; include child-friendly mechanisms for consultation and participation; provide support measures for parents and carers; and ensure effective remedies.⁵⁰ Children’s digital media choices and data control possibilities are shaped by the design and functionalities of communication spaces, control of which rests neither with them, their parents or indeed national regulators.⁵¹

Legal frameworks should cover the full range of unlawful acts which can be committed online,⁵² and there should be a co-regulatory framework that defines the roles and responsibilities of all organisations operating in the digital space.⁵³ Minimum standards that focus on child safety and the full realisation of children’s rights should be established that cover all actors in the chain.⁵⁴

Recommendations

- The Online Safety Codes should take the form of detailed prescriptive code encompassing both protective and preventive measures, prohibiting all forms of violence, exploitation and abuse; include child-friendly mechanisms for consultation and participation; provide support measures for parents and carers; and ensure effective remedies.⁵⁵

⁴⁷ UN Committee on the Rights of the Child, General Comment no 25 (2021) on children’s rights in relation to the digital environment, CRC/C/GC/25, para 38.

⁴⁸ *ibid* para 39.

⁴⁹ *ibid*.

⁵⁰ *ibid* para 73.

⁵¹ Macenaite, M. (2017). From universal towards child-specific protection of the right to privacy online: Dilemmas in the EU General Data Protection Regulation. *New Media & Society*, 19(5), 765–779. <<https://doi.org/10.1177/1461444816686327>> accessed 4 September 2023.


⁵² UN Committee on the Rights of the Child, General Comment no 25 (2021) on children’s rights in relation to the digital environment, CRC/C/GC/25, para 74.

⁵³ 5Rights Foundation, Making Child Online Safety a Reality: Global Toolkit (2022) 185.

⁵⁴ *ibid*.

⁵⁵ *ibid* para 73.

Question 6: How should we design the Code to minimise the potential for conflict and maximise the potential for synergies in how platforms comply with it and the DSA?

 The Online Safety Code needs to be compatible with Ireland's International and Domestic Legislation. Of particular consideration should be the UN Conventions including the Convention on the Rights of the Child, the Irish Constitution and the European Convention on Human Rights.⁵⁶

All public bodies in Ireland, including Coimisiún na Meán, have a responsibility to promote equality, prevent discrimination and protect the human rights of their service users and everyone affected by their policies and plans. This duty, known as the Public Sector Equality and Human Rights Duty, is located in section 42 of the Irish Human Rights and Equality Act 2014.

Recommendations

- Ensure that when designing the Code that particular attention is paid to ensuring it is compliant with international and national human rights law.

⁵⁶ which has been incorporated into Irish Law by the European Convention on Human Rights Act 2003

Question 7: To what extent, if at all, should the Code require VSPS providers to take measures to address content connected to video content?

We have heard from our members that the content connected to video content can often cause significant harm and distress to children and young people, particularly in the context of bullying. At times the video itself may not be the content that is causing harm but when it is considered alongside the content, such as comments connected to the video, it can cause significant distress and harm.⁵⁷ Our members have told us that Travellers and Roma are often targeted in the comments that go with particular videos (for example the poor treatment of animals) which can result in racist content being shared in the comments under the video content.⁵⁸

Recommendation

- Consideration should be given to requiring VSPS Providers to take measures to address content related to video content such as comments etc. This could include requiring VSPS providers to moderate content in comment sections, and have procedures in place for the timely removal of content.

⁵⁷ Children's Rights Alliance member consultation, August 2023.

⁵⁸ Communication received by the Children's Rights Alliance from Pavee Point, 25 August 2023.

Question 8: How should we ask VSPS providers to introduce a feature that allows users to declare when videos contain advertising or other type of commercial communications? Should the Code include specific requirements about the form in which the declaration should take? What current examples are there that you regard as best practice?

User-created video content on social media platforms and video-streaming services (e.g., TikTok, YouTube) frequently involves commercial content and marketing messages. For example, unboxing videos, toy play videos or influencers reviewing products. It can be unclear for children and young people that this content is actually advertising.

The American Academy of Paediatrics has outlined that research on children's understanding of television advertising shows that:

- Children under the age of 8 have 'limited ability to understand the persuasive intent (i.e., that someone else is trying to change their thoughts and behaviour) of the advertiser.'
- Children aged 7 to 11 'can start to recognize television advertising and persuasive intent with their parents' assistance but lack the abstract thinking skills that help individuals recognize advertising as a larger commercial concept.'
- Children and young people over the age of 12 'were able to identify television advertisements (ads) and advertisers' intention to change behaviour'.⁵⁹

The Council of Europe has recommended that 'States should take measures to ensure that children are protected from commercial exploitation in the digital environment, including exposure to age-inappropriate forms of advertising and marketing'.⁶⁰

The UN Committee on the Rights of the Child has reiterated this in their recent General Comment and has recommended that:

'States parties should make the best interests of the child a primary consideration when regulating advertising and marketing addressed to and accessible to children. Sponsorship, product placement and all other forms of commercially driven content should be clearly distinguished from all other content and should not perpetuate gender or racial stereotypes'.⁶¹

Aligned to this, the Committee have recommended that there is a need for the code to ensure that the profiling or targeting of children for commercial purposes is prohibited including practices that 'rely on neuromarketing, emotional analytics, immersive advertising and advertising in virtual and

⁵⁹ The American Academy Of Pediatrics | Policy Statement, July 01 2020, Digital Advertising to Children, <<https://publications.aap.org/pediatrics/article/146/1/e20201681/37013/Digital-Advertising-to-Children?autologincheck=redirected>> accessed 29 August 2023.

⁶⁰ Council of Europe, Guidelines to respect, protect and fulfil the rights of the child in the digital environment (2018) Recommendation CM/Rec(2018)7 of the Committee of Ministers, 20.

⁶¹ UN Committee on the Rights of the Child, General Comment no 25 (2021) on children's rights in relation to the digital environment, CRC/C/GC/25, para 41

augmented reality environments to promote products, applications and services’.⁶² The 2020 WHO-UNICEF-Lancet Commission on the future for the world’s children noted that “commercial marketing of products that are harmful to children represents one of the most underappreciated risks to their health and wellbeing”.⁶³

Digital media advertising has changed dramatically over time and is predicted to account for 60% of global advertising expenditure by 2025.⁶⁴ A 2023 report from UNICEF and the WHO highlights that as marketing communication techniques have moved away from one-size-fits-all spot advertisements towards strategies for fostering engagement, children are now not just passive viewers of commercial messages, but rather ‘active practitioners’ in the commercial communications and marketing.⁶⁵

Recommendations

- The Code should look to ensure that a consistent feature for VSPS providers is introduced across all platforms that places a stringent requirement on users to declare when videos contain advertising and/or commercial communications. It should include a specific requirement for what form the declaration should take. This should be clear, concise, transparent and easy for children and young people to understand.

⁶² UN Committee on the Rights of the Child, General Comment no 25 (2021) on children’s rights in relation to the digital environment, CRC/C/GC/25, para 42.

⁶³ Clark, H., Coll-Seck, A.M., Banerjee, A., Peterson, S., Dalglish, S.L., Ameratunga, S. *et al.* (2020). A future for the world’s children? A WHO–UNICEF–Lancet Commission. *Lancet* 2020; 395: 605–58. <[https://www.thelancet.com/journals/lancet/article/PIIS0140-6736\(19\)32540-1/fulltext#articleInformation](https://www.thelancet.com/journals/lancet/article/PIIS0140-6736(19)32540-1/fulltext#articleInformation)> accessed 4 September 2023.

⁶⁴ WHO, Understanding the digital media ecosystem. How the evolution of the digital marketing ecosystem impacts tobacco, alcohol and unhealthy food marketing (WHO 2022) <<https://apps.who.int/iris/handle/10665/355277>> accessed 4 September 2023.

⁶⁵ UNICEF and WHO, Taking action to protect children from the harmful impact of food marketing: a child rights-based approach. Geneva: World Health Organization and the United Nations Children’s Fund (UNICEF 2023) 7.

Question 9: How should we ask VSPS providers to introduce and design a flagging mechanism in the Code? How can we ensure that VSPS providers introduce the mechanism in a user-friendly and transparent way? How should we ask VSP Providers to report the decisions they've made on content after it has been flagged?

It should not be expected or assumed that a child will be able to identify or report content or conduct which are against a service's community guidelines. The 5Rights Foundation recommend having in place a number of moderation and reporting systems including take down mechanisms and flagging mechanisms.⁶⁶ Currently, there is no consistent flagging system in place for harmful content.

The best interest of the child should be a key focus when considering the design of the flagging mechanism in the code. The Council of Europe (COE) *Guidelines to Respect, Protect and Fulfil the Rights of the Child in the Digital Environment* provide that 'in all actions concerning children in the digital environment, the best interests of the child shall be a primary consideration' and further recommend that States should strike a balance between the child's right to protection and their other rights to freedom of expression, participation and access to information.⁶⁷ The COE also acknowledges the differing levels of maturity and understanding of children at different ages and recommends that States recognise the evolving capacities of children which can mean that the 'policies adopted to fulfil the rights of adolescents may differ significantly from those adopted for younger children'.⁶⁸

An example of how to design a flagging mechanism that responds to the rights of children and young people can be seen in the UK Children's Code regarding the protection of children's data online. The code requires that designated services should provide 'prominent and accessible tools to help children exercise their data protection rights and report concerns.'⁶⁹ The ICO's guidance to services includes that the tools should be prominent and easy for the child to find, age appropriate and easy to use, tailored and specific to the rights they support, and include mechanisms for tracking progress and communicating with the service.⁷⁰ To make tools prominent the ICO suggests services highlight the reporting tools in their set up process and provide a clear icon on the screen display.⁷¹ To make tools age appropriate and easy to use the ICO states that they should be tailored to the age of the child in question.⁷² The ICO provide examples of how to do so in the Code for each age range from 0-5 up to 16-17.⁷³ In order to tailor their tools to support children's rights, the ICO suggest services create a 'download all my data' tool, a 'delete all my data tool' or 'select data for deletion' tool, a 'stop using my data' tool, and a 'correction' tool.⁷⁴ In terms of creating mechanisms that allow parents and children to track the progress of their flagged concern, the ICO state that information should be provided by the service about the timescales for responding to requests and these should be dealt with within the timescales set out at Article 12(3) of the GDPR.⁷⁵ Additionally, in order to

⁶⁶ 5Rights Foundation, 'But how do they know it is a child? Age Assurance in the Digital World'.

⁶⁷ Council of Europe, 'Recommendation CM/Rec(2018)7 of the Committee of Ministers to Member States on Guidelines to respect, protect and fulfil the rights of the child in the digital environment' (COE 2018) <<https://bit.ly/2Xp9hpE>> accessed 26 February 2021, 12.

⁶⁸ *ibid.*

⁶⁹ Information Commissioner's Office, 'Age Appropriate Design: A Code of Practice for Online Services' 8.

⁷⁰ *ibid.* 83-84.

⁷¹ Information Commissioner's Office, 'Age Appropriate Design: A Code of Practice for Online Services' 82.

⁷² *ibid.*

⁷³ Information Commissioner's Office, 'Age Appropriate Design: A Code of Practice for Online Services' 82-84.

⁷⁴ *ibid.* 84.

⁷⁵ *ibid.*

conform with the Code the ICO suggest that services should have mechanisms for children to indicate that they think their complaint or request is urgent, with appropriate prioritisation and the ability to take swift action on ongoing safeguarding issues.⁷⁶ This model could be taken and adapted to specifically relate to video content for the purposes of the Online Safety Code.

The DSA (Article 16) will require platforms to put in place a notification mechanism for illegal content and require them to process the notifications in a timely, diligent, non-arbitrary and objective manner. This should be integrated into the Code being developed. It is important to make the process for flagging content as straightforward and easy to understand for children and young people as possible. Children may find some of the rules set out in community guidelines confusing or struggle to distinguish between what is illegal and what is legal but prohibited by a service.⁷⁷ Requiring users to determine whether they are flagging content under the DSA or the Code would place a significant burden on the user and could act as a deterrent to children and young people flagging illegal and harmful online content.

Recommendations

- Require VSPS to create a consistent flagging system for harmful content and introduce a number of moderation and reporting systems including take down mechanisms.
- The best interest of the child should be a key focus when considering the design of the flagging mechanism in the code.
- The DSA (Article 16) will require platforms to put in place a notification mechanism for illegal content and require them to process the notifications in a timely, diligent, non-arbitrary and objective manner. This should be integrated into the Code being developed.
- Flagging tools should be prominent and easy for the child to find, age appropriate and easy to use, tailored and specific to the rights they support, and include mechanisms for tracking progress and communicating with the service.

⁷⁶ Information Commissioner's Office, 'Age Appropriate Design: A Code of Practice for Online Services' 84.

⁷⁷ Online abuse: teenagers might not report it because they often don't see it as a problem LSE blog by Powell-Jones. May 7th 2019.

Question 10: What requirements should the Code include about age verification and age assurance? What sort of content should be shown by default to users who are logged out or in private browsing mode and whose age cannot be verified or assured? What evidence is there about the effectiveness of age estimation techniques? What current practices do you regard as best practice? Where accounts are not age verified should default privacy settings be used, should content default to universal content and should contact by others be more limited?

The use of age assurance 'is not a silver bullet for keeping children safe online. It is simply a tool to identify that a service is dealing with a child.'⁷⁸ However, age assurance has the potential to drive the 'development of new products and services to create a richer and more diverse digital ecosystem' for children and young people rather than 'being the route to keeping children out of the digital world'.⁷⁹

The principle of data minimisation needs to be central to the design of any age assurance mechanism that is developed. The *Council of Europe Guidelines to respect, protect and fulfil the rights of the child in the digital environment* state that age verification and assurance systems should use methods that are in line with the principle of data minimisation.⁸⁰ The UN Committee on the Rights of the Child has noted that 'digital practices, such as automated data processing, profiling, behavioural targeting, mandatory identity verification, information filtering and mass surveillance are becoming routine. Such practices may lead to arbitrary or unlawful interference with children's right to privacy; they may have adverse consequences on children, which can continue to affect them at later stages of their lives.'⁸¹ Interference with a child's right to privacy should only be permissible if it is 'provided for by law, intended to serve a legitimate purpose, uphold the principle of data minimisation, be proportionate and designed to observe the best interests of the child'.⁸²

There needs to be a range of age assurance solutions developed that can respond to the different situations that children and young people face.⁸³ The 5Rights Foundation have set out that 'many of the changes necessary to make a service age appropriate do not need additional or new age assurance technologies, but rather require services to disable some of their more intrusive or risky design features'.⁸⁴

There is a need to ensure that there are minimum standards put in place for age assurance. This could include 'an explicit risk-based framework that would allow businesses to understand what level of assurance is required in different scenarios.'⁸⁵

⁷⁸ 5Rights Foundation, 'But how do they know it is a child? Age Assurance in the Digital World' 7.

⁷⁹ *ibid* 9.

⁸⁰ Council of Europe, *Guidelines to respect, protect and fulfil the rights of the child in the digital environment* (2018) Recommendation CM/Rec(2018)7 of the Committee of Ministers, 69.

⁸¹ UN Committee on the Rights of the Child, General Comment no 25 (2021) on children's rights in relation to the digital environment, CRC/C/GC/25, para 68.

⁸² *ibid*.

⁸³ 5Rights Foundation, 'But how do they know it is a child? Age Assurance in the Digital World' 7.

⁸⁴ 5Rights Foundation, 'Pathways: How digital design puts children at risk' 11.

⁸⁵ *ibid*.

5Rights Foundation have set out 11 common standards that should inform the development of any age assurance mechanism including:

- Age assurance must be privacy preserving.
- Age assurance should be proportionate to risk and purpose.
- Age assurance should be easy for children to use.
- Age assurance must enhance children’s experiences, not merely restrict them.
- Age assurance providers must offer a high level of security.
- Age assurance providers must offer routes to challenge and redress.
- Age assurance must be accessible and inclusive.
- Age assurance must be transparent and accountable.
- Age assurance should anticipate that children don’t always tell the truth.
- Age assurance must adhere to agreed standards.

Age assurance must be carried out in compliance with children’s rights under National and International law. In order to ensure a rights-based approach to the design and implementation of age assurance measures a human rights analysis should be carried out and measures that are compliant with children’s rights should be adopted. The level of assurance should be proportionate to the nature and level of risk presented by a product or service in relation to the age of the child. It is important that the ‘cumulative nature of risk must also be taken into account, as multiple design features or different parts of a user’s journey combine to create greater risks.’⁸⁶

United Kingdom Children’s Code

The UK Children’s Code regarding the protection of children’s data online, offers a dual option to designated services in order to comply with the standard of age verification or ‘age appropriate application’ as it is termed under the Code. Under the Code designated services are required to take a risk-based approach to recognising the age of individual users either by establishing age ‘with a level of certainty that is appropriate to the risks to the rights and freedoms of children that have arisen from [their] data processing’ or by applying the standards in the Code to all users.⁸⁷ The ICO suggest that Data Protection Impact Assessments (DPIA), which are set out at Standard 2 in the Code, should be used to aid this assessment.⁸⁸ The Code is not prescriptive about the exact methods services should use to establish age or the level of certainty provided focussing instead on the need to use a method that is appropriate to the risk level involved from the service’s data processing.⁸⁹ However, the ICO does provide a non-exhaustive list of options for services to consider including self-declaration, artificial intelligence, third party verification services, account holder confirmation,

⁸⁶ *ibid* 19.

⁸⁷ Information Commissioner’s Office, ‘Age Appropriate Design: A Code of Practice for Online Services’ 32.

⁸⁸ *ibid*.

⁸⁹ Information Commissioner’s Office, ‘Age Appropriate Design: A Code of Practice for Online Services’ 33.

technical measures, and hard identifiers.⁹⁰ The ICO assesses whether a service has chosen the appropriate method by taking account of the products available on the market currently, and in particular for small businesses which cannot develop their own age verification tool due to capacity and resource constraints.⁹¹ Guidance is provided in the Code on how to uphold rights if the collection of personal data is required in order to establish age.⁹² The ICO stresses that while there are tensions between age assurance and compliance with the GDPR, age assurance and the GDPR are compatible if privacy by design solutions are used.⁹³

Recommendations

- Age verification and assurance mechanisms should respect the principle of data minimisation and avoid unlawful or arbitrary interference with the right of the child to privacy.
 - Ensure that any age assurance mechanism introduced is compliant with children's rights under National and International law.
 - There should be a range of age assurance solutions developed which respond to the different situations children and young people face.
 - Ensure that there are minimum standards put in place for age assurance. This could include an explicit risk-based framework that would allow businesses to understand what level of assurance is required in different scenarios.
- Data Protection Impact Assessments and Children's Rights Impact Assessments could be used to monitor the level of interference of age verification mechanisms with the right of the child to privacy and help balance that right with the need for protection online.

⁹⁰ *ibid* 34.

⁹¹ Information Commissioner's Office, 'Age Appropriate Design: A Code of Practice for Online Services' 33.

⁹² *ibid* 35.

⁹³ *ibid*.

Question 11: What requirements should the Code have in relation to content rating? What do you consider to be current best practice? What experiences have you had using content rating systems on platforms and do you think they have been effective? What steps could we ask VSPS to take to ensure content is rated accurately by users?

The UN Committee on the Rights of the Child has recommended that States ‘should encourage providers of digital services used by children to apply concise and intelligible content labelling, for example on the age-appropriateness or trustworthiness of content.’⁹⁴ The Council of Europe has recommended that ‘states should co-operate with a view to promoting standardisation of content classification and advisory labels among countries and across stakeholder groups to define what is appropriate and what is inappropriate for children.’⁹⁵ There are a number of frameworks that could be considered.

A key tool to identify risk and classification of harm is the 4Cs framework. This framework should be considered for adoption in the Online Safety Code.

The CO:RE 4Cs classification recognises that online risks arise when a child:

- Engages with and/or is exposed to potentially harmful content
- Experiences and/or is targeted by potentially harmful contact
- Witnesses, participates in and/or is a victim of potentially harmful conduct
- Is party to and/or exploited by a potentially harmful contract⁹⁶

The 4Cs classification ‘distinguishes between aggressive, sexual and value risks’ along with recognising important cross-cutting risks such as children’s right to privacy and fair treatment.⁹⁷

⁹⁴ UN Committee on the Rights of the Child, General Comment no 25 (2021) on children’s rights in relation to the digital environment, CRC/C/GC/25, para 55.

⁹⁵ Council of Europe, Guidelines to respect, protect and fulfil the rights of the child in the digital environment (2018) Recommendation CM/Rec(2018)7 of the Committee of Ministers, 29, para 121.

⁹⁶ CORE, ‘4 Cs of online risk: Short report & blog on updating the typology of online risks to include content, contact, conduct, contract risk’ <<https://core-evidence.eu/posts/4-cs-of-online-risk>> accessed 28 August 2023.

⁹⁷ *ibid.*

 CORE	Content Child as recipient	Contact Child as participant	Conduct Child as actor	Contract Child as consumer
Aggressive	Violent, gory, graphic, racist, hateful and extremist content	Harassment, stalking, hateful behaviour, unwanted surveillance	Bullying, hateful or hostile peer activity e.g. trolling, exclusion, shaming	Identity theft, fraud, phishing, scams, gambling, blackmail, security risks
Sexual	Pornography (legal and illegal), sexualization of culture, body image norms	Sexual harassment, sexual grooming, generation and sharing of child sexual abuse material	Sexual harassment, non-consensual sexual messages, sexual pressures	Sextortion, trafficking for purposes of sexual exploitation, streaming child sexual abuse
Values	Age-inappropriate user-generated or marketing content, mis/disinformation	Ideological persuasion, radicalization and extremist recruitment	Potentially harmful user communities e.g. self-harm, anti-vaccine, peer pressures	Information filtering, profiling bias, polarisation, persuasive design
Cross-cutting	Privacy and data protection abuses, physical and mental health risks, forms of discrimination			

98

A classification scheme is in place in Australia where the Australian Online Safety Act (2021) defines content as either ‘class 1 material’ or ‘class 2 material’.⁹⁹ Class 1 material and class 2 material are defined by reference to Australia’s National Classification Scheme, which is also used for classification of films, computer games and other publications.

Class 1 material includes material that:

- ‘depicts, expresses or otherwise deals with matters of sex, drug misuse or addiction, crime, cruelty, violence or revolting or abhorrent phenomena in such a way that they offend against the standards of morality, decency and propriety generally accepted by reasonable adults to the extent that they should not be classified
- describes or depicts in a way that is likely to cause offence to a reasonable adult, a person who is, or appears to be, a child under 18 (whether the person is engaged in sexual activity or not), or
- promotes, incites or instructs in matters of crime or violence.’

Class 2 material is material that is, or would likely be, classified as either:

- ‘X18+ (or, in the case of publications, category 2 restricted), or
- R18+ (or, in the case of publications, category 1 restricted) under the National Classification Scheme, because it is considered inappropriate for general public access and/or for children and young people under 18 years old.’

The eSafety Commissioner works with online service providers to ensure access to Class 2 material, which is considered unsuitable for children and young people under 18, is restricted.

⁹⁸ CORE, ‘4 Cs of online risk: Short report & blog on updating the typology of online risks to include content, contact, conduct, contract risk’ <<https://core-evidence.eu/posts/4-cs-of-online-risk>> accessed 28 August 2023.

⁹⁹ Online Safety Act 2021 s106 and s107.

Recommendations

- Content labelling should be concise, intelligible and written in child friendly language.
- Consider adopting the 4C's classification framework for content labelling for child safety online.

Question 12: What requirements should the Code have in relation to parental control features? How can we ensure that VSPS providers introduce the mechanism in a user-friendly and transparent way? Can you point to any existing example of best practice in this area? Should parental controls be ‘turned-on’ by default for accounts of minors or where age is not verified?

While parental controls are one measure for protecting children online they ‘are not a substitute for good design that prioritises user safety’ and can result in parents having a false sense of security ‘while children continue to be exposed to risks due to poor service design’.¹⁰⁰ The most vulnerable children offline are often the most vulnerable online also. Parental control features and safety features have to take this into account as parents may not be in a position to protect their child online.

The Council of Europe has recommended that children’s evolving capacities should be taken into account when businesses establish or update their parental controls.¹⁰¹ Additionally, States should ensure that such controls do not reinforce discriminatory attitudes or infringe on children’s privacy and information rights.¹⁰²

The UK Children’s Code specifies that if a regulated service provides parental controls, they should give the child age appropriate information about this.¹⁰³ If the regulated service allows a parent or carer to monitor their child’s activity online or track their location then they should provide an obvious sign to the child when they are being monitored.¹⁰⁴ ICO ground the basis for this standard within the best interests of the child principle in Article 3 UNCRC, the right of the child to privacy under Article 16, and the requirement under Article 5(1)(a) of the GDPR that any processing of personal data must be lawful, fair and transparent. In terms of conforming to the standard, regulated services should also provide parents with information about the child’s right to privacy and resources for age appropriate discussion between parent and child.¹⁰⁵ The ICO also provide a table with some indicative measures that services could take to ensure compliance that are appropriately targeted at each age group.¹⁰⁶ The Code regulates where a service has parental controls in place but does not require services to have such controls in place.

VSPS must be careful if introducing parental controls to ensure that they do so in a balanced manner that respects the autonomy and privacy rights of the child, cognisant of their developing capacity while also balancing their best interests and safeguarding concerns. The new Online Safety Code should provide guidance on best practice for those services that decide to introduce parental controls.

Recommendations

- Parental Controls should not be a substitute for safety by design features.

¹⁰⁰ https://www.ofcom.org.uk/__data/assets/pdf_file/0027/226269/5rights-foundation.pdf.

¹⁰¹ Council of Europe, Guidelines to respect, protect and fulfil the rights of the child in the digital environment (2018) Recommendation CM/Rec(2018)7 of the Committee of Ministers, 54.

¹⁰² *ibid.*

¹⁰³ Information Commissioner’s Office, ‘Age Appropriate Design: A Code of Practice for Online Services’ 61.

¹⁰⁴ *ibid.*

¹⁰⁵ Information Commissioner’s Office, ‘Age Appropriate Design: A Code of Practice for Online Services’ 62.

¹⁰⁶ *ibid.*

- Where parental controls are adopted, they should respect and reflect the evolving capacities of the child and be compatible with human rights and privacy law.
- Where parental controls are used by a VSPS children and young people who are service users should be given age appropriate and accessible information about this.
- Regulated services should provide parents with information about the child's right to privacy and resources for age appropriate discussion between parent and child.

Question 13: What requirements should the Code contain to ensure that VSPS provide for effective media literacy measures and tools?

Digital and media literacy is an area that requires specific and targeted measures¹⁰⁷ to ensure equal access to the digital environment and the full realisation of children's rights. The UN Committee on the Rights of the Child have stated that parents and guardians should be supported to gain digital literacy in order to support their children in traversing the digital environment in a way which respects their evolving capacities,¹⁰⁸ and educational programmes and materials should be provided in order to develop digital literacy skills.¹⁰⁹ In order to support the full breadth of children's rights, digital literacy education should include both functional and technical competencies, skills related to content creation, and critical thinking around the impacts of the digital environment.¹¹⁰

Digital and media literacy programmes must be accessible to all groups and in particular the most vulnerable. Travellers experience low levels of literacy and low levels of media literacy as a result of exclusion within the education system resulting in low levels of school completion.¹¹¹ Roma experience similarly low levels of media literacy and also face language barriers.¹¹² It is therefore vital that groups such as Travellers and Roma are targeted by VSPS in terms of media literacy measures and tools. It is important that Traveller parents are empowered in relation to parental controls and other controls and tools that may be available to them. This needs to happen in a culturally appropriate way and in consultation with Traveller organisations. Additionally, children who lack resources at home or live in residential care should not be disadvantaged from accessing digital literacy opportunities.¹¹³

Particular efforts should be made to reach those children who have no access to digital technology due to socio-economic or geographic reasons, and those who have access but lack the skills to use or underuse technology due to vulnerability or disability.¹¹⁴ Effective digital literacy should enhance and promote the equality of opportunity and outcomes for all and in particular should promote gender equality by enhancing the use of technology by girls.¹¹⁵ Educational programmes and resources on digital literacy should include information on preventive measures, rights and responsibilities in the digital environment, risk and violation identification, and effective remedies.¹¹⁶ These programmes should enable children to respect fundamental rights, understand what it means to give consent, enable an understanding of what constitutes and how to deal with harmful content including how to seek redress, and to understand the potential consequences of sharing personal information online.¹¹⁷

¹⁰⁷ UN Committee on the Rights of the Child, General Comment no 25 (2021) on children's rights in relation to the digital environment, CRC/C/GC/25, para 11.

¹⁰⁸ *ibid*, para 21.

¹⁰⁹ UN Committee on the Rights of the Child, General Comment no 25 (2021) on children's rights in relation to the digital environment, CRC/C/GC/25, para 32.

¹¹⁰ Council of Europe, Guidelines to respect, protect and fulfil the rights of the child in the digital environment (2018) Recommendation CM/Rec(2018)7 of the Committee of Ministers, 42.

¹¹¹ Pavee Point, Submission to Coimisiún na Meán Call For Inputs on Developing First Online Safety Code 2023.

¹¹² *ibid*.

¹¹³ Council of Europe, Guidelines to respect, protect and fulfil the rights of the child in the digital environment (2018) Recommendation CM/Rec(2018)7 of the Committee of Ministers, 44.

¹¹⁴ *ibid* 45.

¹¹⁵ Council of Europe, Guidelines to respect, protect and fulfil the rights of the child in the digital environment (2018) Recommendation CM/Rec(2018)7 of the Committee of Ministers, 46.

¹¹⁶ *ibid* 48.

¹¹⁷ Council of Europe, Guidelines to respect, protect and fulfil the rights of the child in the digital environment (2018) Recommendation CM/Rec(2018)7 of the Committee of Ministers, 48.

Recommendations

- Digital literacy education should include both functional and technical competencies, skills related to content creation, and critical thinking around the impacts of the digital environment.
- Parents and guardians should be supported to gain digital literacy in order to support their children in traversing the digital environment in a way which respects their evolving capacities.
- Digital literacy programmes must be accessible to all groups and in particular the most vulnerable.
- Educational programmes and resources on digital literacy should include information on preventive measures, rights and responsibilities in the digital environment, risk and violation identification, and effective remedies.

Question 14: How should key aspects of terms and conditions be brought to users' attention? What examples are there of best practice in relation to terms and conditions including content moderation policies and guidelines?

Terms and conditions should be accessible, transparent, fair, and available in child friendly language and recognising that parents and guardians may rely on such terms and conditions as a guide to the suitability of content for their children, businesses should take reasonable steps to ensure they are enforced.¹¹⁸

It is important that published terms:

- use simple language
- aid comprehension
- be concise
- be presented in multiple formats for different age ranges
- be prominent and easy to find
- be presented at the right moments in a user journey
- consider the diverse needs of young people
- not assume adult involvement
- cater for children with accessibility needs
- ensure that consent must be obtained and sought, not assumed
- ensure users are given meaningful choices¹¹⁹

Alongside this the Code should ensure that:

- Terms of agreement should be proportionate to the value young people derive from the service
- Terms of service must be consistently enforced
- Rules must be harmonised and consistent with relevant regulation
- Terms must set out clear rules for what constitutes a breach of terms
- Terms and conditions must clarify what happens when a user makes a complaint¹²⁰

¹¹⁸ Council of Europe, Guidelines to respect, protect and fulfil the rights of the child in the digital environment (2018) Recommendation CM/Rec(2018)7 of the Committee of Ministers, 14, 97.

¹¹⁹ 5Rights Foundation, Tick to Agree Age appropriate presentation of published terms September 202, 10-22.

¹²⁰ ibid 22-32.

In the UK Children’s Code under the transparency standard, the ICO state that if designated services need to draft their terms and conditions in a certain way to be legally robust then they can provide child-friendly explanations in order to meet the standard of providing clear terms, policies and community standards.¹²¹

Recommendations

- Terms and conditions should be accessible, transparent, fair, and available in child friendly language.
- Terms of agreement should be proportionate to the value young people derive from the service.
- Terms of service must be consistently enforced and set out clear rules for what constitutes a breach.
- Terms and conditions should be drafted in child friendly and plain language. If this is not possible services should provide additional child friendly explanations in order to provide clear terms and policies.

¹²¹ Information Commissioner’s Office, ‘Age Appropriate Design: A Code of Practice for Online Services’ 39.

Question 15: How should we ask VSPS providers to address content moderation in the Code? Are there any current practices which you consider to be best practice? How should we address automated content detection and moderation in the Code?

The UN Committee on the Rights of the Child has recommended that ‘Content moderation and content controls should be balanced with the right to protection against violations of children’s other rights, notably their rights to freedom of expression and privacy.’¹²² Further the Committee has stated that State Parties ‘should ensure that digital service providers comply with relevant guidelines, standards and codes and enforce lawful, necessary and proportionate content moderation rules.’¹²³ On automated systems the Committee has recommended that States ‘should ensure that uses of automated processes of information filtering, profiling, marketing and decision-making do not supplant, manipulate or interfere with children’s ability to form and express their opinions in the digital environment.’¹²⁴ The Committee also notes that ‘automated systems may be used to make inferences about a child’s inner state’ and that States should ‘ensure that automated systems or information filtering systems are not used to affect or influence children’s behaviour or emotions or to limit their opportunities or development.’¹²⁵

It is essential that services are not allowed to rely solely on user complaints and are obliged to engage in proactive moderation practices. The 5Rights Foundation have noted that ‘proactive moderation lifts the burden off children to flag and report content and behaviour that violates a service’s community guidelines.’¹²⁶

The Code should ensure that moderation is ‘proportionate to the risk and activities associated with the product or service.’¹²⁷ This would mean that services which are directed at children and young people ‘should pre-moderate all user-generated content’ and services with varied audiences ‘should offer children a higher bar of moderation than other users.’¹²⁸

Moderation must be fair, unbiased and consistent for it to be effective.¹²⁹ The Online Safety Code presents an opportunity for providers to be held to ‘agreed enforceable standards of moderation, including oversight of automated decisions and training and care for human moderators’.¹³⁰

Recommendations

- Content moderation and content controls should be balanced with the right of the child to privacy and freedom of expression.
- Content moderation rules should be necessary and proportionate to the risk and activities associated with VSPS products or services.

¹²² UN Committee on the Rights of the Child, General Comment no 25 (2021) on children’s rights in relation to the digital environment, CRC/C/GC/25, para 58.

¹²³ *ibid.*

¹²⁴ UN Committee on the Rights of the Child, General Comment no 25 (2021) on children’s rights in relation to the digital environment, CRC/C/GC/25, para 61.

¹²⁵ *ibid* para 62.

¹²⁶ 5Rights Foundation, Tick to Agree Age appropriate presentation of published terms September 2021, 34.

¹²⁷ *ibid.*

¹²⁸ 5Rights Foundation, Tick to Agree Age appropriate presentation of published terms September 2021, 34.

¹²⁹ *ibid.*

¹³⁰ 5Rights Foundation, Tick to Agree Age appropriate presentation of published terms September 2021, 35.

- Automated system should not be used to affect or influence children's behaviour or emotions.
- Services directed at children and young people should pre-moderate all user-generated content and services with a varied audience should offer children and young people a higher bar of moderation than other users.
- Moderation should be fair, unbiased, and consistent.

Question 16: What requirements should the Code include about procedures for complaint-handling and resolution, including out-of-court redress or alternative-dispute resolution processes? How frequently should VSPS providers be obliged to report to the Commission on their complaint handling systems and what should those reports contain? Should there be a maximum time-period for VSPS providers to handle user complaints and if so, what should that period be?

The UN Committee on the Rights of the Child recommended in its 2021 General Comment that ‘States parties should ensure that appropriate and effective remedial judicial and non-judicial mechanisms for the violation of children’s rights relating to the digital environment are widely known and readily available to all children and their representatives’.¹³¹ The Committee also recommended that ‘complaint and reporting mechanisms should be free of charge, safe, confidential, responsive, child-friendly and available in accessible formats.’¹³² The Committee is clear that in order to protect children there is a need for complaint handling to be ‘swift to halt any ongoing and future damage.’¹³³

In 2018, the Council of Europe published its Recommendation, *Guidelines to Respect, Protect and Fulfil the Rights of the Child in the Digital Environment* which recommends that States require businesses to meet their responsibilities by requiring them to implement measures and ‘encourage them to co-operate’ with the State and other stakeholders, including children.¹³⁴ It further recommends that Member States should ensure that a child’s right to an effective remedy under the European Convention of Human Rights¹³⁵ is respected and protected when their rights have been infringed online.¹³⁶ This means that States are required to make provision for ‘known, accessible, affordable, and child-friendly avenues through which children, as well as their parents or legal representatives, may submit complaints and seek remedies’.¹³⁷ States and relevant stakeholders such as VSPS should provide children with information in a manner that they can understand on complaints processes and handling so that they are enabled to exercise their participation rights fully.¹³⁸ Guidance is given on what constitutes an effective remedy and it includes:

- inquiry,
- explanation,

¹³¹ UN Committee on the Rights of the Child, General Comment no 25 (2021) on children’s rights in relation to the digital environment, CRC/C/GC/25, para 44.

¹³² *ibid.*

¹³³ UN Committee on the Rights of the Child, General Comment no 25 (2021) on children’s rights in relation to the digital environment, CRC/C/GC/25, para 46.

¹³⁴ Council of Europe, ‘Recommendation CM/Rec(2018)7 of the Committee of Ministers to Member States on Guidelines to respect, protect and fulfil the rights of the child in the digital environment’ (COE 2018) 11.

¹³⁵ European Convention of Human Rights Art 6 and 19.

¹³⁶ Council of Europe, ‘Recommendation CM/Rec(2018)7 of the Committee of Ministers to Member States on Guidelines to respect, protect and fulfil the rights of the child in the digital environment’ (COE 2018) 24.

¹³⁷ *ibid.*

¹³⁸ Council of Europe, ‘Recommendation CM/Rec(2018)7 of the Committee of Ministers to Member States on Guidelines to respect, protect and fulfil the rights of the child in the digital environment’ (COE 2018) 5.

- reply,
- correction,
- proceedings,
- immediate removal of unlawful content,
- apology,
- reinstatement,
- reconnection
- compensation.¹³⁹

Importantly, it provides that the process should be speedy, child-friendly and provide the appropriate redress.¹⁴⁰ In order to be effective it is essential that the Codes provide for a maximum time-period for VSPS providers to handle user complaints that offers and quick and effective resolution for children and young people. The Online Safety Code developed by the Australian eSafety Commissioner states that Tier 1 social media services must resolve complaints within ‘a reasonable time’ and that what constitutes a reasonable time ‘should be based on the scope and urgency of potential harm that is related to a complaint and the source of the complaint.’¹⁴¹

It is important that VSPS providers are required to be transparent in their complaint handling. To this end they should be required to report on their complaint handling systems at a minimum annually.

Recommendations

- Complaint and reporting mechanisms should be free of charge, safe, confidential, responsive, child-friendly and available in accessible formats.
- VSPS should provide children with information in a manner that they can understand on complaints processes and handling.
- The new Online Safety Code should provide for a maximum time-period for VSPS providers to handle user complaints that offers and quick and effective resolution for children and young people and guidance as to what is a reasonable timeframe for responding to complaints.

¹³⁹ *ibid.*

¹⁴⁰ *ibid.*

¹⁴¹ eSafety Commissioner for Australia, Schedule 1 – Social Media Services Online Safety Code (Class 1A and Class 1B Material), 15.

Question 17: What approach do you think the Code should take to ensuring that the safety measures we ask VSPS providers to take are accessible to people with disabilities?

The UN Committee on the Rights of the Child has noted that children with disabilities may be ‘more exposed to risks, including cyberaggression and sexual exploitation and abuse, in the digital environment.’¹⁴² The Committee recommends that states take measures to identify the risks faced by children with disabilities and take steps to ensure they are safe in the digital environment.¹⁴³ This should be done in a way that counters ‘prejudice faced by children with disabilities that might lead to overprotection or exclusion.’¹⁴⁴ It is important that information is provided in accessible formats on safety and protective strategies.¹⁴⁵ One method of ensuring this is equality proofing safety measures and providing guidance on various accessibility methods in place.¹⁴⁶

The Code must respect the evolving capacities of all children including those of children with disabilities or in vulnerable situations.¹⁴⁷ Policies and practices adopted by VSPS under the Code must respect and respond to the needs of these groups in the digital environment and reflect appropriately the differing needs of children of different ages and backgrounds.¹⁴⁸

Recommendations

- Information should be provided in accessible formats on safety and protection strategies.
- Safety measures should be equality proofed as a matter of standard practice.
- Policies and practices adopted by VSPS under the Code must respect and respond to the needs of children and young people with disabilities in the digital environment and reflect appropriately the differing needs of children of different ages and backgrounds.

¹⁴² UN Committee on the Rights of the Child, General Comment no 25 (2021) on children’s rights in relation to the digital environment, CRC/C/GC/25, para 92.

¹⁴³ *ibid.*

¹⁴⁴ UN Committee on the Rights of the Child, General Comment no 25 (2021) on children’s rights in relation to the digital environment, CRC/C/GC/25, para 92.

¹⁴⁵ *ibid.*

¹⁴⁶ For information on equality proofing see: S. Cantillon, K. Lynch, J. Baker, A. Connelly, ‘a Framework for Equality Proofing: A Paper Prepared for the National Economic and Social Forum’ 1995, [A framework for equality proofing: a paper prepared for the national economic and social forum — ResearchOnline \(gcu.ac.uk\)](#).

¹⁴⁷ Council of Europe, Guidelines to respect, protect and fulfil the rights of the child in the digital environment (2018) Recommendation CM/Rec(2018)7 of the Committee of Ministers, 2.

¹⁴⁸ *ibid.*

Question 18: What approach do you think the Code should take to risk assessments and safety by design? Are there any examples you can point us towards which you consider to be best practice?

In 2018, the Council of Europe published its Recommendation, *Guidelines to Respect, Protect and Fulfil the Rights of the Child in the Digital Environment* and noted that the online world is reshaping children's lives in many ways, resulting in 'opportunities for and risks to their well-being and enjoyment of human rights.'¹⁴⁹ Recognising that businesses have a responsibility to respect children's rights,¹⁵⁰ the Council of Europe recommends that States require businesses to meet their responsibilities by compelling them to implement measures and 'encourage them to co-operate' with the State and other stakeholders, including children.¹⁵¹ A key proposal of these Guidelines is that States should require relevant stakeholders to implement safety by design, privacy by design and privacy by default measures, taking into account the best interests of the child.¹⁵² Including these principles in the Online Safety Code would help ensure that, from the planning stages of technology development onward, children are protected. The UN Committee on the Rights of the Child in 2021 recommended that that States should incorporate 'the integration of privacy-by-design into digital products and services that affect children.'¹⁵³

Many of the digital services children and young people use are not designed to protect their rights or meet their needs.¹⁵⁴ Research from the 5Rights Foundation found that 'pathways designed into digital services and products are putting children at risk' with designers tasked with 'optimising products and services for three primary purposes, all geared towards revenue generation.'¹⁵⁵ The Online Safety Code presents a huge opportunity to embed the principle of safety by design into the Irish regulatory framework. It is important that this principle is not incorporated only to services specifically targeted to children and young people but to all the digital services children and young people are likely to actually access.¹⁵⁶

The *Council of Europe Guidelines to respect, protect and fulfil the rights of the child in the digital environment* state that States should require businesses to regularly undertake child-rights impact assessments in relation to digital technologies and demonstrate that they are taking reasonable steps to mitigate risks.¹⁵⁷ Child rights risk assessments should be conducted by business "before their digital products or services could reach or affect children"¹⁵⁸ and businesses should be obliged to "undertake child rights due diligence, which entails that businesses should identify, prevent, and

¹⁴⁹ Council of Europe, 'Recommendation CM/Rec(2018)7 of the Committee of Ministers to Member States on Guidelines to respect, protect and fulfil the rights of the child in the digital environment' (COE 2018) 10.

¹⁵⁰ UN Committee on the Rights of the Child, General Comment No. 16 (2013) on State obligations regarding the impact of the business sector on children's rights CRC/C/GC/16.

¹⁵¹ Council of Europe, 'Recommendation CM/Rec(2018)7 of the Committee of Ministers to Member States on Guidelines to respect, protect and fulfil the rights of the child in the digital environment' (COE 2018) 11.

¹⁵² *ibid* 23.

¹⁵³ UN Committee on the Rights of the Child, General Comment no 25 (2021) on children's rights in relation to the digital environment, CRC/C/GC/25, para 70.

¹⁵⁴ 5Rights Foundation, 'Design of Service' <<https://5rightsfoundation.com/our-work/design-of-service/>> accessed 4 September 2023

¹⁵⁵ 5Rights Foundation, September 2021 Pathways: A Summary Key findings and recommendations from Pathways: How digital design puts children at risk

¹⁵⁶ 5Rights Foundation, 'Design of Service' <<https://5rightsfoundation.com/our-work/design-of-service/>> accessed 4 September 2023

¹⁵⁷ Council of Europe, Guidelines to respect, protect and fulfil the rights of the child in the digital environment (2018) Recommendation CM/Rec(2018)7 of the Committee of Ministers, 95.

¹⁵⁸ The Handbook for policy-makers on the rights of the child in the digital environment by the Council of Europe that accompanies the Recommendation (Livingstone et al., 2020,) 19

mitigate their impact on children’s rights including across their business relationships and within global operations.”¹⁵⁹

Recommendations

- The requirement of safety by design should be one of the key measures included in the Online Safety Code and it should require safety by design to be implemented as standard into all products and services of VSPS.
- Child rights risk assessments should be conducted by VSPS before their digital products or services could reach or affect children.
- VSPS should regularly undertake children’s rights impact assessments in relation to digital technologies and demonstrate that they are taking reasonable steps to mitigate risks.

¹⁵⁹ ibid 72.

Question 19: How do you think that cooperation with other regulators and bodies can help us to implement the Code for VSPS?

Cooperation with other regulators could form an important support for implementation of the Code across key areas of accessibility, human rights compliance and child safety.

In terms of child safety and participation, Tusla could provide an insight on the issues faced by children and young people it works with, and the formal child consultation units in the Department of Children, Equality, Disability, Integration and Youth (DCEDIY) could be coordinated with to ensure proper consultation and engagement from young people on the Code and its implementation.

In terms of human rights compliance and implementation of the public sector duty, the Irish Human Rights and Equality Commission (IHREC) could advise on best practice.

To ensure robust accessibility measures in the development and implementation phases of the Code, the Disability Authority should be coordinated with.

Recommendations

- Cooperate with other public bodies and government departments including Tusla, IHREC, DCEDIY, and the Disability Authority in order to ensure effective implementation of the new Online Safety Code.

Question 21: Do you have any views on how requirements for commercial content arranged by a VSPS provider itself should be reflected in the Code?

Consideration should be given to addressing Harmful Commercial Communications, particularly marketing of high fat, sugar and salt foods and breastmilk substitutes and alcohol.

The American Academy of Paediatrics has outlined that research on children's understanding of television advertising shows that:

- Children under the age of 8 have 'limited ability to understand the persuasive intent (i.e., that someone else is trying to change their thoughts and behaviour) of the advertiser.'
- Children aged 7 to 11 'can start to recognize television advertising and persuasive intent with their parents' assistance but lack the abstract thinking skills that help individuals recognize advertising as a larger commercial concept.'
- Children and young people over the age of 12 'were able to identify television advertisements (ads) and advertisers' intention to change behaviour'.¹⁶⁰

The Council of Europe has recommended that 'States should take measures to ensure that children are protected from commercial exploitation in the digital environment, including exposure to age-inappropriate forms of advertising and marketing.'¹⁶¹

The UN Committee on the Rights of the Child has reiterated this in their recent General Comment and has recommended that:

'States parties should make the best interests of the child a primary consideration when regulating advertising and marketing addressed to and accessible to children. Sponsorship, product placement and all other forms of commercially driven content should be clearly distinguished from all other content and should not perpetuate gender or racial stereotypes.'¹⁶²

Aligned to this, the Committee have recommended that there is a need for the code to ensure that the profiling or targeting of children for commercial purposes is prohibited including practices that 'rely on neuromarketing, emotional analytics, immersive advertising and advertising in virtual and augmented reality environments to promote products, applications and services'.¹⁶³

Recommendations

- Consideration should be given to addressing Harmful Commercial Communications, particularly marketing of high fat, sugar and salt foods, breastmilk substitutes and alcohol.

¹⁶⁰ The American Academy Of Pediatrics | Policy Statement, July 01 2020, Digital Advertising to Children, <<https://publications.aap.org/pediatrics/article/146/1/e20201681/37013/Digital-Advertising-to-Children?autologincheck=redirected>> accessed 29 August 2023.

¹⁶¹ Council of Europe, Guidelines to respect, protect and fulfil the rights of the child in the digital environment (2018) Recommendation CM/Rec(2018)7 of the Committee of Ministers, 20.

¹⁶² UN Committee on the Rights of the Child, General Comment no 25 (2021) on children's rights in relation to the digital environment, CRC/C/GC/25, para 41.

¹⁶³ ibid para 42.

- VSPS should take measures to ensure that children are protected from commercial exploitation in the digital environment, including exposure to age-inappropriate forms of advertising and marketing.
- The best interests of the child should form a primary consideration when regulating advertising and marketing addressed to and accessible to children.
- Sponsorship, product placement and all other forms of commercially driven content should be clearly distinguished from all other content and should not perpetuate gender or racial stereotypes.
- The profiling or targeting of children for commercial purposes should be prohibited.

Question 23: Should the Code have a transition period or transition periods for specific issues? What time frame would be reasonable for a transition period?

It is important that the Online Safety Code comes into force as soon as is possible without delay. Currently platforms are largely unregulated with children and young people experiencing harm online daily. In 2021, CyberSafeKids reported that a quarter of all children have seen or experienced something online in the last year that bothered them, with almost one third of those children having kept it to themselves rather than report it to their parents or someone else.¹⁶⁴

The transition period should be as short as possible to ensure that there is robust protection for children and young people in the digital space. A useful example is the UK Children's Code which provided for a one-year transition period to encourage conformance.¹⁶⁵ For pre-existing services, the Code recommended some measures to take including reviews of processing and pre-existing data protection impact assessments during this period as well as assessing any additional measures that would be needed to conform to the Code.¹⁶⁶ A timeframe like this could be considered.

Recommendations

- It is important that the Online Safety Code comes into force as soon as is possible without delay and the transition period should be as short as possible to ensure that there is robust protection for children and young people in the digital space.

¹⁶⁴ CyberSafeKids, *Annual Report 2021 (2022)* 3.

¹⁶⁵ Information Commissioner's Office, 'Age Appropriate Design: A Code of Practice for Online Services' 21.

¹⁶⁶ *ibid.*



Aighneacht:

**Glaobair ar Ionchuir: Sábháilteacht Ar Líne
An Chéad Chód Ceangailteach ar Líne de chuid na
hÉireann a fhorbairt le haghaidh
Seirbhísí Ardáin Comhroinnte Físeáin**

Meán Fómhair 2023

INTREOIR

Cuireann Conradh na Gaeilge fáilte roimh an deis seo aighneacht a chur isteach don Chéad Chód Ceangailteach ar líne de chuid na hÉireann a fhorbairt le haghaidh Seirbhísí Ardáin Comhroinnte Físeáin.

Is é Conradh na Gaeilge fóram daonlathach phobal na Gaeilge agus saothraíonn an eagraíocht ar son na teanga ar fud na hÉireann uile agus timpeall na cruinne. Is í príomhaidhm na heagraíochta an Ghaeilge a athréimniú mar ghnáth-theanga na hÉireann. Ó bunaíodh é ar 31 Iúil 1893 tá baill an Chonartha gníomhach ag cur chun cinn na Gaeilge i ngach gné de shaol na tíre, ó chúrsaí dlí agus oideachais go fhorbairt meán cumarsáide agus seirbhísí Gaeilge.

Tá Conradh na Gaeilge roghnaithe ag Foras na Gaeilge, an foras uile oileánda ag feidhmiú ar son an dá Rialtas thuaidh agus theas leis an nGaeilge a chur chun cinn, mar cheann de na sé cheanneagraíocht atá maoinithe acu leis an nGaeilge a fhorbairt ar oileán na hÉireann. Go príomha, tá Conradh na Gaeilge roghnaithe le tabhairt faoi chosaint teanga, ionadaíocht agus ardú feasachta ar an Ghaeilge. Tá 180 craobh agus iomaí ball aonair ag Conradh na Gaeilge, agus bíonn baill uile an Chonartha ag saothrú go dian díograiseach chun úsáid na Gaeilge a chur chun cinn ina gceantair féin. Tá breis eolais faoi obair an Chonartha le fáil ag www.cnag.ie.

INTREOIR

Cuireann Conradh na Gaeilge fáilte roimh an deis seo aighneacht a chur isteach maidir leis an gCéad Chód Ceangailteach ar líne de chuid na hÉireann a fhorbairt le haghaidh Seirbhísí Ardáin Comhroinnte Físeáin.

Is deis an Cód Ceangailteach nua seo chun cinntiú go mbeidh cothromas idir an Ghaeilge an Béarla agus cumarsáid á dhéanamh leis an bpobal i dtaobh sábhailteacht ar líne do sheirbhísí ardáin comhroinnte físeáin agus dá réir sin normalú a dhéanamh ar an nGaeilge.

GLAO AR IONCHUIR: SÁBHÁILTEACTH AR LÍNE AN CHÉAD CHÓD CEANGAILTEACH AR LÍNE DE CHUID NA HÉIREANN A FHORBAIRT LE HAGHAIDH SEIRBHÍSÍ ARDÁIN COMHROINNTE FÍSEÁIN

Tá sé luaithe sa cháipéis Glao ar Ionchuir: Sábháilteacht ar Líne – An chéad chód ceangailteach ar líne de chuid na hÉireann a fhorbairt le haghaidh seirbhísí ardáin comhroinnte físeáin ‘Ábhar a spreagann foréigean nó fuath i gcoinne grúpa daoine nó baill de ghrúpa bunaithe ar aon cheann de na forais dá dtagraítear in Airteagal 21 den Chairt um Chearta Bunúsacha an Aontais Eorpaigh. Áirítear leis na forais sin gnéas, cine, dath, bunús eitneach nó sóisialta, gnéithe géiniteacha, **teanga**, reiligiún nó creideamh, tuairim pholaitiúil nó aon tuairim eile, ballraíocht de mhionlach náisiúnta, maoin, breith, míchumas, aois nó claonadh gnéasach.’ (lth. 7).

Mar a fheictear ansin tá teanga san áireamh in Airteagal 21, is gá do Choimisiún na Meán a chinntiú nach bhfuil idirdhealú á dhéanamh ar chúrsaí teanga ná aon fhuathchaint ar an mbonn sin, an Ghaeilge san áireamh. Is maith é, mar sin, go bhfuil ‘teanga’ san áireamh agus molann muid go gcoinneofar san áireamh é.

FORÁIL A DHÉANAMH DO BHEARTA AGUS D’UIRLISÍ ÉIFEACTACHA LITEARTHACHTA NA MEÁN AGUS FEASACHT ÚSÁIDEOIRÍ AR NA BEARTA AGUS NA HUIRLISÍ SIN A ARDÚ.

I 5 j) Ith. 13 deir sé ‘Foráil a dhéanamh do bhearta agus d’uirli sí éifeachtacha litearthachta na meán agus feasacht úsáideoirí ar na bearta agus na huirlisí sin a ardú.’ Ba chóir go mbeidh an litearthacht seo ní amháin a bheith ar fáil i mBéarla ach go mbeidh sé ar fáil i nGaeilge chomh maith.

GNÉ MAIDIR LE CUMARSÁID TRÁCHTÁLA A DHEARBHÚ

Luadh i 5.1.1 (Ith. 13) ‘Gné maidir le Cumarsáid Tráchtála a Dhearbhu – Beart (c)’.

Má tá comhlacht poiblí i mbun cumarsáid tráchtála, is gá a chinntiú go bhfuil an comhlacht sin ag cloí le hait a 6 d’Acht na dTeangacha Oifigiúla (leasú), 2021¹ a deir ‘gur i nGaeilge a bheidh 20 faoin gcéad ar a laghad d’aon fhógraíocht arna déanamh ag an gcomhlacht in aon bhliain’ agus ‘go ndéanfar 5 faoin gcéad ar a laghad d’aon airgead a chaithfidh an comhlacht ar fhógraíocht in aon bhliain a úsáid chun fógraíocht a chur amach i nGaeilge trí na meáin Ghaeilge’. Tá sainmhíniú ar fhógraíocht san acht a deir ‘ciallaíonn fógraíocht ciallaíonn ‘fógraíocht’—

- (a) cumarsáid tráchtála d’aon chineál a bhfuil d’aidhm léi, nó a bhfuil d’éifeacht léi, go díreach nó go neamhdhíreach, táirge nó seirbhís de chuid an chomhlachta poiblí lena mbaineann a chur chun cinn, agus
- (b) cumarsáid d’aon chineál leis an bpobal, i leith na nithe seo a leanas—
 - (i) foireann a earcú,
 - (ii) tionscnaimh reachtaíochta nó bheartais,
 - (iii) talamh nó sócmhainní a cheannach nó a dhíol,
 - (iv) seirbhísí a sholáthar, nó
 - (v) comhchomhairliúchán poiblí;

ciallaíonn “meáin Ghaeilge” aon mheáin ina bhfuil 50 faoin gcéad nó níos mó d’ábhar na meán sin trí mheán na Gaeilge.”’

GNÉ RÁTÁLA ÁBHAIR

Tá sé ráite i 5.1.4 (Ith. 16) ‘Gné Rátála Ábhair – Beart (g)’. Ba chóir a chinntiú go mbeidh aon fhógraí a bhaineann le gné rátála ábhair ar fáil go dátheangach. Tá na siombail ann faoi láthair dátheangach² (seachas an ceann do PG, ag seasamh do Parental Guidance), seo deis anois chun cinntiú go mbeidh gach fógra/siombal go hiomlán dátheangach.

LITEARTHACHT SNA MEÁIN

Ba chóir a aithint go bhfuil pobal a úsáideann Gaeilge taobh istigh agus taobh amuigh den Ghaeltacht. Mar sin má tá tábhacht faoi leith ann go mbeidh tuiscint an phobail ar ábhar a fhoilsítear i meáin chlóite, chraolta, ar líne nó meáin eile, is gá go mbeidh na bearta agus d’uirli sí éifeachtacha litearthachta sna meáin ar fáil i nGaeilge chomh maith leis an mBéarla.

¹ <https://www.irishstatutebook.ie/eli/2021/act/49/enacted/ga/print#sec6>

² <https://www.ifco.ie/en/ifco/pages/guidelines>

LÁIMHSEÁIL GEARÁIN

Ba chóir go mbeidh Coimisiún na Meán in ann déileáil le gearáin i mBéarla nó i nGaeilge, le gearáin maidir leis an bhfuathchaint a bhaineann le Gaeilge san áireamh (breis eolais ar fáil ag an nasc thíos)³. Is gá cloí le hAirteagal 21 de Chairt um Chearta Bunúsacha an Aontais Eorpaigh agus Coimisiún na Meán ag plé leis na gearáin seo.

CONCLÚID

Ba chóir anailís sochtheangeolaíochta a dhéanamh agus an cód seo a dhréachtú, le cinntiú go mbeidh an códchleachtais ag teacht leis na rudaí seo a leanas:

- Airteagal 21 de Chairt um Chearta Bunúsacha an Aontais Eorpaigh
- Acht na dTeangacha Oifigiúla (Leasú), 2021
- Gné Rátála Ábhair
- Bearta agus Uirlisí Éifeachtacha Litearthacht sna Meáin
- Láimhseáil Gearáin

Tá Conradh na Gaeilge ar fáil má tá aon cheist maidir le haon ghné den aighneacht seo.

³ https://peig.ie/wp-content/uploads/2019/12/13NOLL2019_Aighneacht_Reachta%C3%ADocht_Fuathchaint.pdf



**Coimisiún na Meán Call for Inputs
on an Online Safety Code for Video-Sharing Platform Services**

**Submission by the Ombudsman for Children's Office
4 September 2023**

1. Introduction	1
2. Priorities and Objectives	1
3. Online Harms	7
4. Measures to be taken by VSPS providers	8
• Age verification	8
• Parental controls.....	9
• Media literacy	11
• User complaints	11

1. Introduction

Following its establishment in March 2023, Coimisiún na Meán (Commission) announced on 11 July that it is seeking views through a Call for Inputs on developing its first online safety code, which will apply to video-sharing platform services (VSPS).¹

The OCO welcomes the Commission's decision to gather views from the public on how the Commission should develop the code and the opportunity this consultation presents for us to provide an initial input at an early stage in the process. The OCO also welcomes the Commission's decision to undertake a phased process to gather input from children on the code.

The OCO is an independent statutory body, which was established in 2004 under the Ombudsman for Children Act 2002 (2002 Act). Under the 2002 Act, as amended, the Ombudsman for Children has two core statutory functions:

- to promote the rights and welfare of children up to the age of 18 years, and
- to examine and investigate complaints made by or on behalf of children about the administrative actions of public bodies, schools and voluntary hospitals that have or may have adversely affected a child.

The OCO has prepared this submission pursuant to section 7(4) of the 2002 Act, which provides that the Ombudsman for Children may advise on any matter relating to the rights and welfare of children.

In preparing this submission, the OCO is mindful that the submissions made in response to this Call for Inputs will assist the Commission in its task of information-gathering and reflection as it begins to draft the code. Accordingly, the overall aim of this submission is to set out the OCO's preliminary, high-level observations on several questions raised by the Commission in its Call for Inputs that we believe merit consideration by the Commission in completing this task. The OCO understands that the new code for VSPS will consider online safety in respect of adults and children. However, given the OCO's statutory remit under the 2002 Act, our submission focuses on children.

From the OCO's perspective, the Commission's work to develop Ireland's first online safety code presents a significant opportunity for the Commission, as a newly established independent statutory body, to situate the code within a human rights framework, to place service users, and particularly children, at the centre of the code, and, in doing so, to set an important precedent as regards adopting and promoting a human rights-based approach to the regulation of online safety.

2. Priorities and Objectives

Question 1 of the Call for Inputs asks about the main priorities and objectives of the first binding code for VSPS.² The Call for Inputs states that the Commission will take a child-centred approach to developing the code where it impacts children.³ In this regard, it refers to Article 24 of the EU Charter of Fundamental Rights (Charter) and Article 3 of the UN Convention on the Rights of the Child (CRC).

¹ Coimisiún na Meán, [Coimisiún na Meán seeks views for developing Ireland's First binding Online Safety code](#), 11 July 2023; Coimisiún na Meán (2023), [Call For Inputs: Online Safety](#).

² Coimisiún na Meán (2023), [Call For Inputs: Online Safety](#), p. 9.

³ *Ibid.*, p. 5.

The OCO welcomes that the Commission intends to take a child-centred approach to developing the code and that the Call for Inputs refers in particular to international and EU children's rights standards. In this regard, and as the Commission may be aware, following its most recent periodic review of Ireland's implementation of the CRC in January 2023, the UN Committee on the Rights of the Child (Committee) recommended that the State ensure that the Online Safety Commissioner pays particular attention to the protection of children who fall under its mandate, in line with children's rights standards.⁴

As the Commission knows, having ratified the CRC in 1992, Ireland has an obligation under international law to respect, protect and fulfil the rights set out in the CRC for all children in the State.

Among the CRC rights that are engaged in the online environment are:

- children's right to freedom of expression, which includes the right to seek, receive and impart information and ideas (Article 13)
- children's right to freedom of thought, conscience and religion (Article 14)
- children's right to freedom of association and peaceful assembly (Article 15)
- children's right to protection from arbitrary or unlawful interference with privacy, family, home or correspondence (Article 16)
- children's right to access information and materials from a variety of sources and to be protected from harmful information (Article 17)
- children's right to be protected from all forms of violence, abuse and exploitation (Articles 19, 34 and 36)
- children's right to the highest attainable standard of health (Article 24)
- children's right to education (Articles 28 and 29), and
- children's right to engage in play and recreational activities and to participate freely in cultural life and the arts (Article 31).

Four CRC rights are recognised as integral to the realisation of all children's rights set out in the CRC. These four general principles are:

- children's right to non-discrimination (Article 2)
- children's right to have their best interests treated as a primary consideration in all matters affecting them (Article 3)
- children's right to life, survival and development (Article 6), and
- children's right to express their views freely in all matters affecting them and to have due weight given to their views, in accordance with their age and maturity (Article 12).

The Committee states that in all decisions, measures or actions concerning children, the State should adopt a child rights-based approach, which entails respecting the child as a rights-bearing person and which is best achieved by furthering the realisation of all of the rights set out in the CRC.⁵

The CRC places an obligation on the State, as the primary duty bearer, to respect, protect and fulfil these rights. This obligation extends to the impact on the rights of children of the activities of business enterprises that are operating in the jurisdiction of the State. In 2013, the Committee

⁴ UN Committee on the Rights of the Child (2023), [Concluding observations on the combined fifth and sixth periodic reports of Ireland](#), CRC/C/IRL/CO/5-6, para. 22.

⁵ UN Committee on the Rights of the Child (2011), [General comment No. 13 \(2011\) The right of the child to freedom from all forms of violence](#), CRC/C/GC/13, para. 59.

published a general comment setting out guidance on the State's obligations in this regard. In this general comment, the Committee sets out three types of obligations placed on the State with respect to business enterprises:

- Respect: ensure that all private actors within the jurisdiction respect children's rights,
- Protect: prevent business enterprises from causing or contributing to abuses of children's rights, and
- Fulfil: create an environment in which children's rights can be fully realised.⁶

In setting out a framework for States' implementation of their obligations to children in this area, the Committee highlights the importance of legislative, regulatory and enforcement measures. In particular, the Committee advises that States must provide stable, clear and predictable legal and regulatory environments, which enable business enterprises to respect children's rights.⁷

In 2021, the Committee published a general comment on children's rights in the digital environment, which provides guidance to States on how they can respect, protect and fulfil children's rights online. The Committee reiterates that States should take measures, including through the development, monitoring, implementation and evaluation of legislation, regulatory frameworks and codes, to ensure compliance by businesses with their obligations to prevent their online services from being used in ways that cause or contribute to violations or abuses of children's rights and to provide children and their parents with prompt and effective remedies.⁸

The importance of taking a child rights-based approach to regulation is also highlighted at European level by the Council of Europe and the European Union. In 2018, the Committee of Ministers of the Council of Europe adopted a recommendation to Member States on Guidelines to respect, protect and fulfil the rights of the child in the digital environment. Having regard to the rights of the child set out under the CRC, the recommendation recommends that governments of Member States require business enterprises to meet their responsibility to respect the rights of the child in the digital environment. The recommendation includes guidance on the development of national legal frameworks that apply to businesses operating in the digital environment and recommends that States should create a clear and predictable legal and regulatory environment, which helps businesses and other stakeholders meet their responsibility to respect the rights of the child in the digital environment through their operations.⁹

EU law applicable to VSPS providers equally places emphasis on ensuring respect for children's rights. As the Commission is aware, the Charter applies to EU Member States when implementing EU law and Article 24 of the Charter reiterates key principles set out in the CRC, including that:

- the best interests of the child must be a primary consideration in all actions relating to children, whether taken by public authorities or private institutions
- children have a right to protection and care as is necessary for their wellbeing, and
- children may express their views freely and such views shall be taken into consideration on matters which concern them in accordance with their age and maturity.

⁶ UN Committee on the Rights of the Child (2013), [General comment No. 16 \(2013\) on State obligations regarding the impact of the business sector on children's rights](#), CRC/C/GC/16, paras. 26-29.

⁷ Ibid., para. 29 and para. 53.

⁸ UN Committee on the Rights of the Child (2021), [General comment No. 25 \(2021\) on children's rights in relation to the digital environment](#), CRC/C/GC/25, paras. 35-39.

⁹ Council of Europe (2018), [Guidelines to respect, protect and fulfil the rights of the child in the digital environment: Recommendation CM/Rec\(2018\)7 of the Committee of Ministers](#), para. 78.

The Audiovisual Media Services (AVMS) Directive states that the AVMS Directive respects fundamental rights and observes the principles recognised by the Charter and that it seeks to promote the application of the rights of the child enshrined in the Charter.¹⁰ It also states that EU Member States must carefully balance the rights set out in the Charter, including the rights of the child, when taking appropriate measures to protect children from harmful content.¹¹ Similar aims are set out in the Digital Services Act (DSA), which states that the DSA should be interpreted and applied in accordance with the fundamental rights set out in the Charter.¹²

Under the Online Safety and Media Regulation Act 2022, one of the functions of the Commission is to ensure that the interests of children are protected.¹³ When preparing an online safety code, the Commission is required to have regard in particular to levels of risk of harm, and particularly harm to children, from the availability of harmful online content or exposure to it, as well as the rights of users of online services, which may include children.¹⁴

Having regard to the Commission’s status and functions as an independent statutory body, together with the Commission’s obligations as a State actor under international and European law to uphold children’s rights and ensure that ICT service providers respect children’s rights, the OCO encourages the Commission to situate Ireland’s first online safety code within a human rights framework and to do so in a manner that has specific and explicit regard to children’s rights. In this way, and having regard to children, the Commission can both require and support VSPS providers to take a child-centred, rights-based approach to the design, development, delivery, monitoring and review of their services.

The Committee states that a child rights-based approach requires the adoption of an approach that is guided at all times by the four general principles of the CRC.¹⁵ In particular, the Committee states that the four general principles of the CRC should serve as a guide for determining the measures needed to guarantee the realisation of children’s rights in the digital environment.¹⁶

The OCO suggests that one way in which the Commission could mobilise the first online safety code to promote a rights-based approach by VSPS providers could be by specifying a set of cross-cutting, rights-based principles in the code. Allowing for the fact that the code will cover adults and children, such guiding principles could include, but not be limited to, the four general principles of the CRC. The inclusion of such guiding principles would serve to demonstrate an expectation on the part of the Commission that VSPS providers must respect human rights, including children’s rights. It could also facilitate the Commission to monitor VSPS providers’ compliance with the code from a rights perspective.

¹⁰ Directive (EU) 2018/1808 of the European Parliament and of the Council of 14 November 2018 amending Directive 2010/13/EU on the coordination of certain provisions laid down by law, regulation or administrative action in Member States concerning the provision of audiovisual media services (Audiovisual Media Services Directive) in view of changing market realities, preambular paragraph 60.

¹¹ Ibid., preambular paragraph 51.

¹² Regulation (EU) 2022/2065 of the European Parliament and of the Council of 19 October 2022 on a Single Market For Digital Services and amending Directive 2000/31/EC (Digital Services Act), preambular paragraphs 40 and 153 and Article 1(1).

¹³ Online Safety and Media Regulation Act 2022, s 7(2)(b).

¹⁴ Ibid., ss 139M(f)-(g).

¹⁵ UN Committee on the Rights of the Child (2011), [General comment No. 13 \(2011\) The right of the child to freedom from all forms of violence](#), CRC/C/GC/13, para. 59.

¹⁶ UN Committee on the Rights of the Child (2021), [General comment No. 25 \(2021\) on children’s rights in relation to the digital environment](#), CRC/C/GC/25, para. 8.

We note that such an approach would be consistent with the principles-led approach taken in codes previously adopted by the Broadcasting Authority of Ireland.¹⁷ It would also be consistent with codes relevant to online safety that have been adopted in some other countries, such as in the UK,¹⁸ which take approaches that include embedding children's rights as one principle among a wider set of principles to guide implementation of the code or aligning a code's guiding principles with the rights and principles set out in the CRC.

Given that the scope of the code will be broader than children, the Commission might give consideration to other relevant international human rights instruments in determining the other guiding principles to include in the code. Such standards include the UN Guiding Principles on Business and Human Rights adopted by the UN Human Rights Council in 2011.¹⁹ As with the guidance provided to States by the Committee, the UN Guiding Principles on Business and Human Rights include the principle that business enterprises should respect human rights as a global standard of expected conduct. In order to meet this responsibility, the UN Guiding Principles state that business enterprises should have in place a policy commitment to meet their responsibility to respect human rights, a human rights due diligence process to identify, prevent, mitigate and account for how they address impacts on human rights, and processes to enable remediation of any adverse human rights impacts.

In a report published in 2018, the UN Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression (Special Rapporteur) refers to the UN Guiding Principles in stating that human rights standards provide a framework for holding both States and companies accountable to users.²⁰ The Special Rapporteur sets out the human rights principles that should guide online content regulation. These include: human rights by default; due diligence; transparency; accountability; remediation; legality; necessity and proportionality; and non-discrimination.²¹

The OCO encourages the Commission to seriously consider grounding Ireland's first online safety code in cross-cutting, rights-based principles, which incorporate core child rights principles.

A further measure that the Commission could take to both oblige and support VSPS providers to adopt a child rights-based approach in respect of children is to require them to implement child rights due diligence. Like broader human rights due diligence obligations set out in frameworks such as the UN Guiding Principles on Business and Human Rights, the Committee states that businesses should be required to undertake child rights due diligence in order to meet their obligation to respect children's rights.²² This requires a process of child rights impact assessment (CRIA) to be undertaken by business enterprises.²³

¹⁷ Broadcasting Authority of Ireland, [Codes & Standards](#).

¹⁸ Home Office (2020), [Interim Code of Practice on Online Child Sexual Exploitation and Abuse](#).

¹⁹ Office of the High Commissioner for Human Rights (2011), [Guiding Principles on Business and Human Rights](#).

²⁰ Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression (2018), [Report of the Special Rapporteur to the Human Rights Council on online content regulation](#), A/HRC/38/35, paras. 41-48.

²¹ Ibid.

²² UN Committee on the Rights of the Child (2013), [General comment No. 16 \(2013\) on State obligations regarding the impact of the business sector on children's rights](#), CRC/C/GC/16, para. 62; UN Committee on the Rights of the Child (2021), [General comment No. 25 \(2021\) on children's rights in relation to the digital environment](#), CRC/C/GC/25, para. 38. See also: Council of Europe (2018), [Guidelines to respect, protect and fulfil the rights of the child in the digital environment: Recommendation CM/Rec\(2018\)7 of the Committee of Ministers](#), paras. 94-95.

²³ UN Committee on the Rights of the Child (2021), [General comment No. 25 \(2021\) on children's rights in relation to the digital environment](#), CRC/C/GC/25, para. 38; Council of Europe (2018), [Guidelines to respect, protect and fulfil the rights of the child in the digital environment: Recommendation CM/Rec\(2018\)7 of the Committee of Ministers](#), paras. 94-95.

CRIA is identified by the Committee as a key measure to implement children's rights and involves examination of the potential impacts of laws, policies, decisions or services on children and the enjoyment of their rights and identification of ways to prevent or mitigate any negative impacts.²⁴ At a minimum, the CRC, including its general principles, should be used as a framework for conducting CRIA and CRIA should have special regard for any differentiated impact of measures to be taken on children.²⁵ It is notable that, following its review of Ireland's combined fifth and sixth reports on the implementation of the CRC, the Committee recommended that the State introduce mandatory requirements for the business sector to undertake assessments of, consultations on and full public disclosure of the children's rights impacts of their business activities and their plans to address such impacts.²⁶

The OCO notes that the provisions of the 2022 Act concerning online safety codes place an emphasis on assessing, preventing and mitigating risk, with particular reference to risks of harm to children. Section 139K provides that an online safety code may make provision to ensure that service providers take appropriate measures to minimise the availability of harmful content online and risks arising from the availability of and exposure to such content.²⁷ It also states that an online safety code may provide for the assessment by service providers of the availability of harmful online content on services, of the risk of it being available, and of the risk posed to users by harmful online content.²⁸ When preparing an online safety code, the 2022 Act requires the Commission to have regard to particular matters, including the levels of risk of exposure to harmful online content when using designated online services and the levels of risk of harm, and in particular harm to children, from the availability of harmful online content or exposure to it.²⁹

The need to assess and mitigate risks to fundamental rights, including the rights of the child, is also reflected in the provisions of the DSA. As noted in the Commission's Call for Inputs, the DSA requires providers of very large online platforms and very large online search engines to assess the systemic risks of their services and take appropriate mitigating measures in observance of fundamental rights.³⁰ Included among the four categories of systemic risk that such providers are required to assess is the actual or foreseeable negative effect on the exercise of fundamental rights in the Charter, including the rights of the child.³¹ Measures that such providers must take to mitigate identified risks may include targeted measures to protect the rights of the child.³²

The OCO therefore welcomes that the Commission's Call for Inputs suggests that the code could require VSPS providers to carry out bespoke risk assessments of harmful content and that such

²⁴ UN Committee on the Rights of the Child (2013), [*General comment No. 14 \(2013\) on the right of the child to have his or her best interests taken as a primary consideration \(art. 3, para. 1\)*](#), CRC/C/GC/14, para. 99.

²⁵ See also: Council of Europe (2020), [*Handbook for policy makers on the rights of the child in the digital environment to support the implementation of Recommendation CM/Rec\(2018\)7 of the Committee of Ministers of the Council of Europe on Guidelines to respect, protect and fulfil the rights of the child in the digital environment*](#); Digital Futures Commission (2021), [*Child Rights Impact Assessment: A tool to realise children's rights in the digital environment*](#).

²⁶ UN Committee on the Rights of the Child (2023), [*Concluding observations on the combined fifth and sixth periodic reports of Ireland*](#), CRC/C/IRL/CO/5-6, para. 13(b).

²⁷ Online Safety and Media Regulation Act 2022, s 139K(2)(a).

²⁸ *Ibid.*, s 139K(4)(c).

²⁹ *Ibid.*, s 139M(e)-(f).

³⁰ Regulation (EU) 2022/2065 of the European Parliament and of the Council of 19 October 2022 on a Single Market For Digital Services and amending Directive 2000/31/EC (Digital Services Act), preambular paragraphs 79 and Article 34(1)(b).

³¹ *Ibid.*, preambular paragraph 80.

³² *Ibid.* Article 35(1)(j).

assessment could include a child rights impact assessment.³³ We also welcome the focus on assessing risks and identifying mitigation measures in the design and development of services.³⁴

The OCO encourages the Commission to consider specifying CRIA in the first online safety code as an approach to implementing requirements associated with identifying, preventing and mitigating risks of harm to children and their rights.

3. Online Harms

Question 1 of the Call for Inputs asks about the main online harms that the code should address and why.³⁵ The Commission states that reference to online harms in the Call for Inputs includes harm caused by harmful online content, illegal content, inappropriate content and commercial communications collectively.

The OCO notes that the Call for Inputs states that the Commission intends the code to complete the transposition of Article 28b of the AVMS Directive into Irish law, in line with the Commission's duty to develop a code in this regard under section 139K(3) of the 2022 Act. The Commission states that it also needs to consider how to use its code-making powers to address wider categories of online harm that are set out in the 2022 Act. Beyond the obligation to transpose Article 28b into Irish law, section 139K of the 2022 Act provides the Commission with the discretion to make codes that ensure protections are taken by online services against harmful online content set out in section 139A of the 2022 Act, which includes the offence-specific categories of online content and other categories of online content. In addition, the Commission envisages that the code will complement the DSA, when it comes into effect in February 2024,³⁶ and later asks stakeholders how the code can be designed to minimise conflict and maximise synergies in how platforms comply with the DSA.³⁷

Given that VSPS providers will have obligations under the AVMS Directive, the 2022 Act, and the DSA, and that there is some overlap between the categories of online harm covered in these three instruments, an optimal approach might be for the first online safety code to cover VSPS providers' obligations in respect of all types of harm. **Having regard to the international children's rights guidance that States should ensure a clear and predictable regulatory environment for service providers and that regulations for service providers should be comprehensive and effective in ensuring protection of children from harmful content and risks online, the OCO encourages the Commission to give consideration to covering all relevant harms applicable to VSPS providers in one code.**

In this regard, the OCO also notes that the Commission states in the Call for Inputs that it presumes it will adopt one code for VSPS providers, at least initially.³⁸ If it is not feasible for the Commission to prepare a code that addresses VSPS providers' obligations across domestic and EU law, **the OCO suggests that an alternative approach might be to focus the initial code on those areas where there is alignment between the 2022 Act, the AVMS Directive and the DSA. If such an approach was to provide regulatory clarity and coherence that can support compliance by VSPS providers, it could serve the interests of service users, including children.**

³³ Coimisiún na Meán (2023), [Call For Inputs: Online Safety](#), pp. 22-23.

³⁴ Ibid.

³⁵ Ibid., p. 9.

³⁶ Ibid., p. 5.

³⁷ Ibid., p. 11.

³⁸ Ibid., p. 9.

4. Measures to be taken by VSPS providers

- **Age verification**

Question 10 of the Commission's Call for Inputs asks stakeholders about the requirements that should be included in the code in respect of age verification.

Article 28(3)(f) of the AVMS Directive includes age verification among the measures that Member States should require VSPS providers to take, as appropriate, with respect to content that may impair children's physical, mental or moral development. The DSA also includes age verification among a list of risk mitigation measures that VLOPs or VLOSEs may take to protect the rights of the child.³⁹

The OCO welcomes that the Commission plans to include a requirement that VSPS providers introduce appropriate age-verification mechanisms to protect children from online harms in the code. The Committee states that robust age verification systems should be used to prevent children from access to illegal products or services and such systems should be consistent with data protection and safeguarding requirements.⁴⁰ The Council of Europe similarly recommends that effective systems of age verification are used to ensure protection against access to content or services that are legally restricted with reference to specific ages, using methods consistent with the principle of data minimisation.⁴¹

The OCO is aware that efforts to introduce standards for age assurance are underway at international and European levels. The International Organization for Standardization (ISO) has prepared a Working Draft Age Assurance Systems Standard to provide a common framework for age assurance.⁴² At EU level, the euCONSENT project aims to develop EU-wide infrastructure to enable online age verification and parental consent, in consultation with children, academic experts, NGOs and other stakeholders in child rights and online protection.⁴³

At a national level, the 5Rights Foundation in the UK has outlined a set of common child-centred standards that should apply to age assurance.⁴⁴ These include that age assurance:

- must be privacy preserving
- should be proportionate to risk and purpose
- should be easy for the child to use
- must enhance children's experiences, not merely restrict them
- must offer a high level of security
- must offer routes to challenge and redress
- must be accessible and inclusive
- must be transparent and accountable

³⁹ Regulation (EU) 2022/2065 of the European Parliament and of the Council of 19 October 2022 on a Single Market For Digital Services and amending Directive 2000/31/EC (Digital Services Act), Article 35(1)(j).

⁴⁰ UN Committee on the Rights of the Child (2021), [General comment No. 25 \(2021\) on children's rights in relation to the digital environment](#), CRC/C/GC/25, para. 114.

⁴¹ Council of Europe (2018), [Guidelines to respect, protect and fulfil the rights of the child in the digital environment: Recommendation CM/Rec\(2018\)7 of the Committee of Ministers](#), para. 56.

⁴² ISO, [ISO/IEC WD 27566](#).

⁴³ <https://euconsent.eu/>.

⁴⁴ 5Rights Foundation (2021), [But how do they know it is a child? Age Assurance in the Digital World](#).

- should anticipate that children don't always tell the truth
- must be subject to agreed standards
- must be rights-respecting.

In Ireland, the Data Protection Commission's Fundamentals for a Child-Oriented Approach to Data Processing sets out a non-exhaustive list of criteria for a risk-based approach to age verification that should be considered by organisations who decide to implement age verification mechanisms.⁴⁵ These include:

- the type of data being processed
- the sensitivity of personal data being processed
- type of service offered to the child
- accessibility of personal data collected to other persons
- the further processing of personal data.

From the OCO's perspective, it is vital that online service providers enforce age restrictions appropriately and effectively and that the onus is on the service provider to ensure that no child below the minimum age to use their service or access content on their service can do so. **The OCO encourages the Commission to ensure that the requirements set out in the code in relation to age verification provide for VSPS providers to respect children's rights and to do so in a way that has regard to and balances different children's rights online appropriately.**

- **Parental controls**

Under Question 12, the Commission asks about the requirements that the code should contain in relation to parental control features.

Article 28(3)(h) of the AVMS Directive includes parental controls among the appropriate measures that Member States should require VSPS providers to take, as appropriate, to protect children from content that may impair their physical, mental or moral development. The preamble of the AVMS Directive suggests that effective parental controls are among the strictest measures, which should be applied to the most harmful content.⁴⁶ While not defined, parental controls may include placing restrictions on the time that children can spend online, placing restrictions on the content that a child can access or share, placing restrictions on the activities that a child can engage in, or monitoring of children's online activities.⁴⁷

Parents and caregivers play an important role in providing assistance to children in exercising their rights online. Under Article 5 of the CRC, parents are recognised as having the primary responsibility for the upbringing and development of the child and as having the best interests of the child as their basic concern. Under Article 18 of the CRC, States undertake to respect the responsibilities, rights and duties of parents to provide, in a manner consistent with the evolving capacities of the child, appropriate direction and guidance to children in the exercise of their rights, and to render appropriate assistance to parents in doing so. The Committee defines the evolving capacities of the

⁴⁵ Data Protection Commission (2021), [Children Front and Centre: Fundamentals for a Child-Oriented Approach to Data Processing](#), pp. 47-48.

⁴⁶ Directive (EU) 2018/1808 of the European Parliament and of the Council of 14 November 2018 amending Directive 2010/13/EU on the coordination of certain provisions laid down by law, regulation or administrative action in Member States concerning the provision of audiovisual media services (Audiovisual Media Services Directive) in view of changing market realities, [20].

⁴⁷ B. Zaman and M. Nouwen (2016), [Parental controls: advice for parents, researchers and industry](#).

child as an enabling principle that addresses the process of maturation and learning through which children progressively acquire competencies, understanding and increasing levels of agency to take responsibility and exercise their rights.⁴⁸

In this regard, parental controls may offer one means by which parents can reduce the risk of a child's exposure to online harm. While acknowledging the role that parental controls may play as part of a range of measures to enable parents to engage with their children in preventing risk, research has highlighted however that the use of parental controls may also impede children's exercise of their other rights online.⁴⁹ Indeed, international children's rights standards and guidance emphasise the need to ensure that a balance of children's rights is achieved by ICT providers when developing parental control measures.

The Committee notes that monitoring or surveillance of children's online activities presents problems for respecting children's right to privacy.⁵⁰ In particular, parental controls, if not implemented carefully, may prevent a child from accessing a helpline or searching for sensitive information. It states that parents' monitoring of a child's digital activity should therefore be proportionate and in accordance with the child's evolving capacities. Similarly, the Council of Europe notes that such controls should be developed and deployed taking into account children's evolving capacities and their rights to non-discrimination, privacy and access to information, in accordance with their age and maturity.⁵¹ The Committee states that in seeking to provide an appropriate balance between respect for the evolving capacities of adolescents and appropriate levels of protection, consideration should be given to a range of factors affecting decision-making, including the level of risk involved, the potential for exploitation, understanding of adolescent development, recognition that competence and understanding do not necessarily develop equally across all fields at the same pace and recognition of individual experience and capacity.⁵²

The Committee also emphasises the important role of the State in providing assistance to parents to give appropriate direction and guidance to children when online. It states that States should raise awareness among parents of the need to respect children's evolving capacities and privacy and support parents in acquiring knowledge of the risks to children to help them assist children in the realisation of their rights. This guidance should support parents to achieve an appropriate balance between protecting the child and respecting their emerging autonomy.⁵³

The OCO encourages the Commission to include a requirement in the code that, where a VSPS provider intends to develop and deploy parental control measures on its service, such controls should be applied in such a way that respects children's evolving capacities, having regard to international children's rights standards and guidance. The OCO further suggests that the code could require VSPS providers to provide associated guidance for parents on the proportionate use of parental controls, taking into account children's rights and evolving capacities.

⁴⁸ UN Committee on the Rights of the Child (2016), [General comment No. 20 \(2016\) on the implementation of the rights of the child during adolescence](#), CRC/C/GC/20, para. 18.

⁴⁹ B. Zaman and M. Nouwen (2016), [Parental controls: advice for parents, researchers and industry](#).

⁵⁰ UN Committee on the Rights of the Child (2021), [General comment No. 25 \(2021\) on children's rights in relation to the digital environment](#), CRC/C/GC/25, para. 76.

⁵¹ Council of Europe (2018), [Guidelines to respect, protect and fulfil the rights of the child in the digital environment: Recommendation CM/Rec\(2018\)7 of the Committee of Ministers](#), para. 54.

⁵² UN Committee on the Rights of the Child (2016), [General comment No. 20 \(2016\) on the implementation of the rights of the child during adolescence](#), CRC/C/GC/20, para. 20.

⁵³ UN Committee on the Rights of the Child (2021), [General comment No. 25 \(2021\) on children's rights in relation to the digital environment](#), CRC/C/GC/25, para. 86.

- **Media literacy**

Under Question 13, the Commission asks about the requirements that the code should contain to ensure that VSPS providers provide for effective media literacy measures and tools.

Article 28b(3)(j) of the AVMS Directive includes effective media literacy measures and tools, and raising users' awareness of those measures and tools, among the list of appropriate measures that VSPS providers should be required to adopt, as appropriate. The Commission states that it intends for the code to implement this measure, including to ensure that users of VSPS understand the features, systems and procedures put in place by VSPS providers to protect citizens from online harms. In this regard, the Commission asks about the requirements that the Code should contain to ensure that VSPS providers provide for effective media literacy measures and tools.

Media literacy, including provision of child-friendly information to children and of information to parents on the measures available on online platforms to protect children on the services they use, can ensure that children are supported to exercise their rights online as well as deal with associated risks to their right to protection from harm. In line with children's right to seek and receive information under Article 13 of the CRC, the Committee and the Council of Europe state that States should encourage ICT providers to provide public, easily accessible, child-friendly and age-appropriate information and educational materials to children and parents in line with children's evolving capacities and in a language that they understand, in order to support children's safe and beneficial digital activities.⁵⁴ This includes information on matters such as a providers' terms of service, unacceptable behaviours and appropriate remedies (including on how and to whom to make a complaint), reporting mechanisms, and how to request help and counselling.

The OCO encourages the Commission to make it a requirement in the code that VSPS providers should ensure that child-friendly information on the measures put in place by VSPS providers to protect children from harmful content online and to respond to harmful content when using the service is made available, easily accessible and presented in multiple formats to children and their parents/guardians.

- **User complaints**

Question 16 of the Call for Inputs addresses the handling by VSPS providers of user complaints. The Commission states that it expects the code to require VSPS providers to establish and operate transparent, easy-to-use and effective procedures for handling users' complaints and to report to the Commission at regular intervals on the handling of communications from users.

The OCO welcomes the proposal to include requirements relating to complaints-handling in this code. The Commission provides examples of instances in which people using a VSPS may wish to make a complaint to a VSPS provider. This could include children who may wish to make a complaint about the actions taken by a VSPS provider, such as a complaint about a content moderation decision made about content that the child uploaded to a VSPS, or a complaint about the way in

⁵⁴ UN Committee on the Rights of the Child (2021), [General comment No. 25 \(2021\) on children's rights in relation to the digital environment](#), CRC/C/GC/25, para. 36, para. 39 and para. 55; Council of Europe (2018), [Guidelines to respect, protect and fulfil the rights of the child in the digital environment: Recommendation CM/Rec\(2018\)7 of the Committee of Ministers](#), para. 20, para. 59 and para. 68.

which a VSPS provider responded to a report that the child made about alleged harmful content available on the VSPS.

As the Commission knows, children can face particular challenges in accessing and participating in complaints processes that affect them. Complaints procedures and practices therefore need to be adapted to meet children's specific needs.⁵⁵ The Committee has stated that States should ensure that businesses provide effective complaint mechanisms for children when their rights have been abused in the digital environment.⁵⁶ Such complaint mechanisms should be free of charge, safe, confidential, responsible, child-friendly and available in accessible formats to all children, their parents and their representatives.⁵⁷ The Committee also states that remedial mechanisms should take into account the vulnerability of children and the need to be swift to halt ongoing and future damage.⁵⁸ The Council of Europe has similarly stated that States should ensure the provision of available, known, accessible, affordable, and child-friendly avenues through which children, as well as their parents or legal representatives, may submit complaints and seek remedies.⁵⁹ States should ensure children are provided with guidance on how and to whom to make a complaint and parents or carers should also be informed of such mechanisms and appropriate remedies. Mechanisms should ensure that access to remedies is speedy and child-friendly and provides appropriate redress to children.⁶⁰

Informed by our experience of dealing with complaints in the context of discharging our statutory complaints function, the OCO published a Guide to Child-Centred Complaints Handling in 2018.⁶¹ The purpose of the guide is to encourage and support organisations, which provide services to children and make decisions that impact on children, to deal with complaints in accordance with good practice and in a child-centred manner. The Guide sets out seven core principles of good practice for dealing with complaints by or on behalf of children, as well as measures that can be taken to translate these principles into practice:

- openness and accessibility,
- best interests of the child,
- participation of children,
- transparency and communications,
- timeliness,
- fairness, and
- monitoring and review.

In particular, the Guide encourages organisations to:

- provide any particular supports that children or their representatives may need during the complaints process,
- involve children in the development of information materials about the complaints process
- seek the views of the child affected by the complaint and address any barriers that may exist for children in expressing their views freely, and

⁵⁵ Ombudsman for Children's Office (2018), [A Guide to Child-Centred Complaints Handling](#).

⁵⁶ UN Committee on the Rights of the Child (2021), [General comment No. 25 \(2021\) on children's rights in relation to the digital environment](#), CRC/C/GC/25, para. 48.

⁵⁷ Ibid., para. 44.

⁵⁸ Ibid., para. 46.

⁵⁹ Council of Europe (2018), [Guidelines to respect, protect and fulfil the rights of the child in the digital environment: Recommendation CM/Rec\(2018\)7 of the Committee of Ministers](#), para. 67.

⁶⁰ Ibid., para. 68.

⁶¹ Ombudsman for Children's Office (2018), [A Guide to Child-Centred Complaints Handling](#).

- seek feedback from children as part of a regular review of the complaints policy and procedures in place.

Having regard to the above, **the OCO encourages the Commission to consider including a requirement in the code that VSPS providers must put in place a child-friendly complaints process, which facilitates complaints to be made by as well as on behalf of children using their service.**



CYBERSAFE KIDS

Response to Coimisiún na Meán's Call For Inputs

Developing Ireland's First Binding Online Safety Code for Video-Sharing Platform Services

3. Online Harms

3.1 What online harms should the Code address?

Question 1: What do you think our main priorities and objectives should be in the first binding Online Safety Code for VSPS? What are the main online harms you would like to see it address and why?

3.1.1: Algorithmic Recommendations:

It will be extremely important to address the way that **recommendation algorithms** are used in all of the popular VSPS, specifically in relation to child users. It has been proven in various pieces of research (two of which are cited in the footnotes) that the algorithm, whilst initially based around interests/viewing history etc will rapidly ‘up the ante’ for increased engagement, regardless of whether or not the user is a child, showing increasingly harmful content over a short period of time.¹ While TikTok has announced that European users will be able to turn off personalisation functions, the For You and the Live feeds as well as TikTok search, will show popular videos, not algorithmically recommended videos, based on past user behaviour and interests. In the interests of not profiling children or indeed in not permitting automated decisions to be made as regards child users, we believe that this function should be employed by VSPS for child users.

3.1.2: Age Verification

We note the provisions of the AVSMD at 28.3b, in so far as: VSPS are to establish and operating age verification systems for users of VSPS with respect to content which may impair the physical, mental or moral development of minors. We believe the Binding Online Safety Code is an opportunity to prescribe the age at which these measures should apply. The age of digital consent Ireland is 16. We note that recent judgment has provided that consent may be used from hereon in terms of processing social media users’ personal data. Hence the digital age of consent for minor may become more to the fore in terms of VSPS and other such platforms. Often the minimum age of users is prescribed by the organisation on internal rules of service, at 13. However, our research indicates that 37% of children we surveyed between the ages of 8-12 were on TikTok and 76% of respondents had a YouTube account, despite the fact that 13 is the minimum age requirement to own a YouTube account.²

The Digital Services Act provides age verification measures must be put in place.

We see that in other European jurisdictions (for example France) the legislature is building age verification requirements into national legislation. We believe that VSPS being commercial entities will

¹ YouTube Leads Young Gamers to Videos of Guns, School Shootings’ Tech Transparency Project [2023], Source: <https://www.techtransparencyproject.org/articles/youtube-leads-young-gamers-to-videos-of-guns-school> and 5Rights Foundation in partnership with Revealing Reality: Pathways: How digital design puts children at risk [July 2021], Source: <https://5rightsfoundation.com/in-action/new-research-shows-children-directly-targeted-with-graphic-content-within-as-little-as-24-hours-of-creating-an-online-social-media-account.html>

² CyberSafeKids Annual Report 2023, (September 2023) source: https://www.cybersafekids.ie/wp-content/uploads/2023/08/CSK_Data-Trends-Report-2023-Sept-5.23-.pdf



not act voluntarily but rather will require the regulator to entice and compel profit yielding organisations to comply with these requirements. The technology exists to ensure accurate age verification measures can be employed across platforms, but the will appears to be lacking in the industry.

Age verification measures are becoming mandatory in jurisdictions throughout the globe. While we acknowledge that much of the work in this area appears to be in the data protections sphere, with the Age-Appropriate Design Code in the UK and COPPA2 and the Californian Age-Appropriate Design Code, all making recent headlines, the clear fact remains that VSPS possess the technology to ensure that their users are over a certain age threshold. Whether this technology is then employed in order to meet compliance under a data protection regime and/or under online safety codes, this technology can and should be utilised. We believe that unless age verification is set down in a prescribed form in the within codes, that there will have been a missed opportunity to carve out an appropriate age threshold for application across VSPs uniformly. Any age verification measures however, should still protect the anonymity of the child (i.e. appropriate third party providers are preferred). Moreover, robust technological measures will not be applied uniformly across VSPS without the creation of a mandatory requirement to do so.

3.1.3. Age Assurance

There should be a focus on making certain that online services that attract children are using robust age-assurance measures to ensure that younger users will have safer and more age-appropriate experiences on their platforms, including any content recommendations.

According to our latest trends and usage data, over a quarter (26%) of children (8-16yrs) have seen or experienced something online in the last year that “bothered” them.³ ‘Bothered’ was defined in the question as something that ‘upset them, scared them or made them wish they hadn’t seen it’.

Online Harms

In our experience, the main online harms to be addressed are as follows:

- **Age-inappropriate content for children**, including:
 - **Pornography** (according to Commonsense Media, the average at which a child first sees pornography is 12) and 15% of children surveyed had seen it by the age of 10.⁴ We know from research by Childline and by CARI that exposures to pornography at a young age can have devastating consequences both for the viewer but also that incidents of peer on peer sexual harm increase in populations where children are exposed to pornography. We know that 1 in 5 children surveyed had come across sexual material online: here 18% of children aged 9 and up had come across sexual

³ CyberSafeKids, 2022/3 Academic Year Trends and Usage data [2023], source: https://www.cybersafekids.ie/wp-content/uploads/2023/08/CSK_Data-Trends-Report-2023-Sept-5.23-.pdf

⁴ Commonsense Media: 'Teens and Pornography' report [2022]: Source: <https://www.common sense media.org/research/teens-and-pornography>

content online.⁵ We know from a recent Court decision Ireland which made international headlines, the Judge commented:

“One disturbing element is that you (the accused) have been watching pornography from the age of 11. I think it is truly shocking that this is available to vulnerable, impressionable young people. Clearly these companies are making vast sums of money from selling pornographic material.

“More rigorous restrictions should be imposed on them to prevent this harmful material being available to young children.”⁶

- A recent study by the Children’s Commissioner for England revealed that sexual violence commonly seen in pornography was found in half of police interview transcripts of child-on-child sex abuse cases. ⁷ European states have taken steps to require the pornography industry to block online access to minors, with France being the leading example:⁸ This is an opportune time for Irish Regulators to ensure that the Safety Codes are prescriptive in terms of requiring VSPs to ensure minors are not exposed to adult content while on their platforms.
- **Extreme violence, horror and torture** in this category. We were recently contacted by a very concerned parent whose child had viewed a video of a cat being tortured and she asked us, “what am I supposed to do now? He can’t unsee it and is really distressed”. It’s worth bearing in mind that a proportion of young children (28% of 8 - 12 year old boys according to our latest Trends and Usage data) are playing over-18s games. The most cited over-18s games by those surveyed were Grand Theft Auto and Call of Duty. Both games are known for adult content, including sex and violence.
- **Self-harm/suicide content** – note the Molly Russell case in the UK and the conclusions of the Coroner’s inquest, which found that: “Molly Rose Russell died from an act of self-harm whilst suffering from depression and the negative effects of on-line content”.⁹
- **Pro-anorexia/eating disorder content**
- **Hate Speech** defined as to include any kind of communication in speech, writing or behaviour that attacks or uses pejorative or discriminatory language with reference to a person or a group on the basis of their inherent/ protected characteristics – in other words, based on their religion, ethnicity, nationality, race, colour, ancestry, gender or other identity factor. Includes dehumanization, which targets individuals or groups by calling them subhuman, comparing them to animals, insects, pests, disease or any other non-human entity.¹⁰ Hate Speech is potentially going to be subject to

⁵ National Advisory Council for Online Safety Report of a National Survey of Children, their Parents and Adults regarding Online Safety 2021 204409_b9ab5dbd-8fdc-4f97-abfc-a88afb2f6e6f (2).pdf

⁶ “Boy, 13, who sexually assaulted student in Cork had been watching porn since he was 11” Irish Examiner, 19 May 2023, available at : <https://www.irishexaminer.com/news/courtandcrime/arid-41143108.html>

⁷ Evidence on Pornography’s Influence on Harmful Sexual Behaviour Between Children , Children’s Commissioner for England <https://lnkd.in/gAX28cWG>

⁸ <https://lnkd.in/gnKpB6FB>

⁹ Molly Russell - Prevention of future deaths report - 2022-0315, source: https://www.judiciary.uk/wp-content/uploads/2022/10/Molly-Russell-Prevention-of-future-deaths-report-2022-0315_Published.pdf

¹⁰ [WEF Typology of Online Harms 2023.pdf \(weforum.org\)](https://www.weforum.org/publications/2023/01/01/WEF-Typology-of-Online-Harms-2023.pdf)

legislation,¹¹ however until this comes into being, it appears that the necessity to ensure the swift removal of hate speech from platforms is gathering pace.

- **Inappropriate contact by online predators:** There have been some disturbing trends identified during the pandemic with some of the key agencies involved in monitoring (including INTERPOL, the National Center for Missing & Exploited Children (NCMEC) in the US, the Internet Watch Foundation (IWF) in the UK and Hotline.ie in Ireland) all reporting significant increases in child sexual abuse material (CSAM) found online in recent years. NCMEC's 2022 CyberTipline report noted an increase of proliferation of CSAM of 9% between 2021 and 2022 (almost 32 million reports), but an overall increase of 47% since 2020.¹² An even more alarming increase was seen in attempts to contact a child for sexual exploitation and grooming purposes (online enticement) - an 82% increase between 2021 and 2022 (80,524 report in 2022).¹³
- **Cyberbullying:** over the past academic year, almost two thirds (62%) of teachers in Ireland dealt with online safety incidents, including cyberbullying more than once in their school over the past year - 21% reported dealing with 5+ incidents in that timeframe.¹⁴ 25% of Irish children aged 8-12 and 40% of children aged 12-16 reported to us that they had experienced cyberbullying over the past year.¹⁵

Question 2: What types of online harms do you think should attract the most stringent risk mitigation measures by VSPs? How could we evaluate the impact of different types of harms e.g. severity, speed at which harm may be caused? Is there a way of classifying harmful content that you consider it would be useful for us to use?

- We appreciate that this is categorised as illegal content, rather than harmful content, but the removal of CSAM should be top priority as well as much stronger preventative measures taken to stop contact between children and online predators as well on online platforms, as the ability to upload such material. In line with the European Commission's plans under Commissioner Ylva Johansson, the electronic service providers (ESPs) should be compelled to scan their services for such material and to remove it as quickly as possible, as well as carry out regular risk assessments on their services. ESP reports do make up the vast majority of reports to the CyberTipline so many of them are active in this area. Some of the ESPs are reporting significantly higher numbers but it is not necessarily indicative of where most of the material is hosted or the size of the user base, but more to do with the sophistication of the tools being used to detect it. Meta Inc, for example, provided over 80% of the reports on its collective services (26 million of the overall 31 million) in 2022.¹⁶ NCMEC said in a statement at the time of publication of the 2021 report "*Higher numbers of reports can be indicative of a variety of things including larger numbers of users on a platform or how robust an ESP's efforts*

¹¹ Criminal Justice (Incitement to Violence or Hatred and Hate Offences) Bill 2022

¹² NCMEC CyberTipline report 2022, source: <https://www.missingkids.org/cybertiplinedata>

¹³ Ibid.

¹⁴ CyberSafeKids 2023

¹⁵ Ibid.

¹⁶ NCMEC CyberTipline 2022

are to identify and remove abusive content. NCMEC applauds ESPs that make identifying and reporting this content a priority and encourages all companies to increase their reporting to NCMEC. These reports are critical to helping remove children from harmful situations and to stopping further victimization”.¹⁷ They also noted in their latest report that the reports they get are “just the tip of the iceberg” as regards CSAM on the internet.

- In terms of harmful content, such as content being used to victimise or bully, the key focus should be on **timely intervention**, especially if it relates to a child user. This is why we urge the CNM to **prioritise putting in place the individual complaints mechanism (ICM)**. We have already reported separately to the Online Safety Commissioner examples of cases where there were incredibly slow response times from the VSPS providers. VSPS providers should be given very clear timeframes within which they should respond to user complaints as part of the safety standard to which they must adhere (24 - 48 hours). We can cite cases where there were very slow response times or no response at all. We would also urge the ICM to have stringent timelines attached to it - for the initial triage of the case but also for any takedown notices issued, as is the case with the E-Safety Commission in Australia. Complaints should be timestamped and sent as a notification of the user so timelines can be closely monitored in the event that there is an unsatisfactory response.
- As well as a focus on the timeliness of the response from ESPs to user complaints and reports, **there also needs to be a focus on quality outcomes especially if a case relates to a child, with the child’s needs always at the centre**. Where content is distressing a child but does not reach the thresholds imposed internally by VSPS in terms of internal content rules or community standards, a mechanism for complaints and reports received by or about children should be categorised as priority and classified as a different category and threshold, including review of any comments attached to the relevant content that show an intent to bully or humiliate.
 - *We had a case where the VSPS provider did not agree with a mother that the content she wished to get removed on behalf of her son was harmful as it did not violate their community standards. The content (7 short videos) did not seem harmful at first view (and as a consequence were not removed). And they yet had been the cause of a horrible bullying campaign against this boy by his school peers because they perceived him to appear “babyish” in them. This campaign eventually forced him to change schools. A teacher in his new school reached out to us because the boy remained terrified that his peers in his new school would find those videos and that it would start again. We contacted the VSPS provider on his behalf and were able to get them removed, not on the basis that they violated the community standards, but because he was 10 when he posted them - technically below the age at which he should have had an account.*

¹⁷ Tillman, R. (2022, March 19). Reports of online child exploitation increased 35% in 2021. NY1. Retrieved January 6, 2023, from <https://www.ny1.com/nyc/all-boroughs/news/2022/03/18/national-center-missing-exploited-children-online-reports-increase>



Question 3: Do you have reports, academic studies or other relevant independent research that would support your views? If you do, please share them with us with links to relevant reports, studies or research.

Cybersafe Kids Year in Review 2021-2022 , September 2022, available here: [Trends and Usage Report Academic Year 2022/23](#)

Data Protection Commission Children Front and Centre: Fundamentals for a Child-Oriented Approach to Data Processing (2021)

World Economic Forum, Toolkit for Digital Safety Design Interventions and Innovations: Typology of Online Harms August 2023, available at: [WEF_Typology_of_Online_Harms_2023.pdf](#) (weforum.org)

See footnotes and references to cases included.

4. Overall Approach to the Code

4.1 How prescriptive or flexible should the Code be?

It is our view that the Code should be both detailed and prescriptive for the following reasons:

- We believe that a flexible code is akin to self-regulation and we know that this does not work because these measures will be contrary to their natural inclination towards commercial benefit for the companies. It is why legislation was brought in, in the first place.
- This Code will be the basis on which penalties will be imposed so they need to be very specific, clear and prescriptive. The Companies need to be clear when they are breaching them, otherwise they won't know how to adhere it.
- The process could allow flexibility in order avoid being cumbersome (i.e. by allowing time and space for remedial action within a specified timeframe before incurring a penalty) but the Code should not – it must be clear and specific.
- To reiterate a point made above the Code needs to be clear, specific and prescriptive regarding timelines for takedown notices in relation to complaints handling.
- We believe that it is worth differentiating out within the Code, approaches to dealing with harms to children as opposed to adults. There are specific harms to children that will require clear and timely responses both from the companies but also, if needed, from the OSC through the Individual Complaints Mechanism.
- The above point also relates to attempts to directly contact children by adults who wish to groom or extort them.
- We referenced above the fact that children are at times being fed age-inappropriate content. We believe the onus should be on the VSPS' to restrict access to age-inappropriate content to children and that they should be held accountable for any age-inappropriate content that does reach them. This will require robust age-assurance measures being in place, which should be part of the Code.

The Code could specify in detail the measures we expect VSPS providers to take to address online harms.

Question 4:**4.3 How should the Code take account of the Digital Services Act (“DSA”)?**

Given that we are likely to be regulating for Europe for the DSA because we host the majority of the VLOPS, we feel that it is important to focus time, energy and resources into the areas that are complementary between the OSMR Act and the DSA. By working in a complementary way, we would be taking advantage of this overlap, allowing for more robust oversight. There should be no scope for conflict.

4.4 How should the Code address content connected to video content?

The comments and shares related to any video in question are inherently part of the offending content. It feeds the likes and the shares and can, in some circumstances, be the evidence of the bullying/harm (i.e. if the content of the video content is benign but the likes and comments create the harm).

It is important to define in the Code what amounts to connected video content (i.e. the likes, shares, comments) so that it can adequately addressed through the Code.

5. Measures to be taken by Video-Sharing Platforms

Terms and Conditions used by VSPS are part of their *internal* regulatory procedure while the Codes will be focused on *external* regulation. We note that in many cases, the T&Cs may not be applicable to real world examples. For example, we note that most of the VSPS’ have clear T&Cs relating to minimum age requirements for their service and yet this is regularly and consistently undermined by the numbers of underage users on their services (note CSK Trends and Usage data 2022/23). For that reason, prescriptive binding Codes would provide a mandate on VSPS to bring into practice coherent and real-world measures to provide safety to children online, including investing in available technologies to ensure they know the age of the child user on their platforms and, from there, to ensure that those child users are not algorithmically fed age inappropriate content and/or harmful content.

There should be very clear expectations in relation to what is included in the VSPS T&Cs and how they are written – i.e. they should be written in a way that is clear and understandable by children. An analogy could be draw from the transparency requirements under GDPR and the distillation of complex data processing into child friendly infographics, words and pictures to allow a child to be fully informed as to how their data are processed. Similarly, the distillation of VSPS community standards and T&Cs into child friendly format should comprise part of the measures to be taken by VSPS.



The Code will need to require VSPS to provide clearly written T&Cs for their services and community guidelines. These must be written in a clear, intelligible and child-friendly manner and should be easy to find on the service in question.

The Codes should address content comprising of 'pranks' or challenges. Significantly, when teens and older children are targets, there has been loss of life.

In recent challenges parents have involved their children, for example, the "egg crack challenge" sees parents breaking eggs on their child or toddler's heads.¹⁸ Often such content creates huge traffic for the VSPS but results in harm to the subject. Codes should provide for certain criteria in circumstances where these challenges are resulting in harm or likely to result in harm. For example, a flagging system where the hashtag it uses, or a prompt where content is being uploaded.

5.1 Online Safety Features for Users

Complaints Handling (already considered above)

In addition, we believe that report buttons often go unanswered for weeks at a time all the while the offending post remains online. We believe that a time record should be created to allow the child to keep a record of the report- as such a **Timestamp/notification** should be created for when a report/complaint is submitted so the user and VSPS has clear record of when such a report was submitted. This is an essential means of measuring timebound responses.

This is also important to inform compliance with requirements under the OSMR Act where the complainant is obliged to exhaust measures with the OSP before escalating complaints to the individual complaints mechanism in due course.

5.1.1 Feature for Declaring Commercial Communications – Measure (c)

Targeted advertising to children should be tackled under the Code and provision should be made for children to not be targeted advertising or profiled. This should be prohibited anyway, under data protection legislation but continues to happen.

The Code should also take into account the reliance on child models and child promoters on influencer platforms. Guidance around this should issue to parents but in particular where a child is promoting brands there should be a prompt or warning from the VSPS. While parents are often the party promoting this content, the VSPS benefit from engaged traffic around popular posts, often parenting posts.

5.1.3 Age Verification and Age Assurance Features – Measure (f)

Please see our comments above.

¹⁸CTA news report, Is the new TikTok 'Egg Crack Challenge' all it's cracked up to be? (Aug 2023), source: <https://www.youtube.com/watch?v=v69PXufcp24>



We are of the view that the use of age assurance measures on the part of the VSPS is essential, but we acknowledge that private browsing presents challenges. We aren't sure how best to address these challenges.

One consideration could be that there is a default age for private browsing of 12 years old. The VSPS would have to be compelled to do this obviously via the Code. Another consideration would be that a child can access vital support services if needed from a private browser– i.e. LGBTQI information from reputable bodies. Such support services should not be age-gated.

Question 10: What requirements should the Code include about age verification and age assurance? What sort of content should be shown by default to users who are logged out or in private browsing mode and whose age cannot be verified or assured? What evidence is there about the effectiveness of age estimation techniques? What current practices do you regard as best practice? Where accounts are not age verified should default privacy settings be used, should content default to universal content and should contact by others be more limited?

5.1.4 Content Rating Feature – Measure (g)

We fully support the idea of classification frameworks for content/websites.

Question 11: What requirements should the Code have in relation to content rating? What do you consider to be current best practice? What experiences have you had using content rating systems on platforms and do you think they have been effective? What steps could we ask VSPS to take to ensure content is rated accurately by users?

Question 12: What requirements should the Code have in relation to parental control features? How can we ensure that VSPS providers introduce the mechanism in a user-friendly and transparent way? Can you point to any existing example of best practice in this area? Should parental controls be 'turned-on' by default for accounts of minors or where age is not verified?

Question 13: What requirements should the Code contain to ensure that VSPS provide for effective media literacy measures and tools?

We know that existing age-restrictions do not work. From our own research, 84% of 8-12 year olds have their own social media and/or instant messaging account, despite minimum age restrictions of at least 13 on all of the popular services. 28% of 8-12 year old boys are playing over-18s games such as Grand Theft Auto and Call of Duty.¹⁹

CyberSafeKids provides talks and resources to parents and educators around parental controls and digital media literacy. We fully support any reference to these measures in the Code but feel that real-world education and campaigns, similar to what we and others provide, are vital to support these regulatory measures.

¹⁹ CSK Trends & Usage data 2023



Parental controls are not a silver bullet. They can support child safety online but there is still considerable responsibility on parents as well as on VSPS to ensure that children will be safer on using their services. Too often, in our experience, these companies point to their parental controls (thereby putting the onus on parents) instead of investing in more substantive child safety infrastructure.

Question 14: How should we ask VSPS providers to address online harms in their terms and conditions in the Code, including the harms addressed under Article 28b? How should key aspects of terms and conditions be brought to users' attention? What examples are there of best practice in relation to terms and conditions including content moderation policies and guidelines?

As above, child friendly explainers should accompany by T&Cs

5.2.2 Applying Terms and Conditions (Content moderation decisions) – Measures (a) & (b)

For certain categories of content, such as incitement to violence, hatred and to cyberbullying, delays can compound the damage, we suggest that this type of content should be removed pending the review and/or decision. We have seen cases where delays have had a detrimental effect, as illustrated in this article.²⁰

Question 15: How should we ask VSPS providers to address content moderation in the Code? Are there any current practices which you consider to be best practice? How should we address automated content detection and moderation in the Code?

5.2.3 Complaint Handling – Measure (i)

Question 16: What requirements should the Code include about procedures for complaint-handling and resolution, including out-of- court redress or alternative-dispute resolution processes? To what extent should these requirements align with similar requirements in the DSA? What current practices could be regarded as best practice? How frequently should VSPS providers be obliged to report to the Commission on their complaint handling systems and what should those reports contain? Should there be a maximum time-period for VSPS providers to handle user complaints and if so, what should that period be?

We believe that VSPS providers should have an acceleration channel available for child users and their guardians to submit complaints.

As noted above, prescriptive timeframes in relation to the handling of complaints, are essential.

5.3 Possible Additional Measures and Other Matters

5.3.1 Accessible Online Safety Features

²⁰ Newton, Casey, 'The unbearable slowness of Meta's oversight board' (Aug 2023) source: https://open.substack.com/pub/platformer/p/the-unbearable-slowness-of-metas?r=5pgwv&utm_campaign=post&utm_medium=email

The Code should be prescriptive in relation to how the interface is structured and presented to users so that all VSPS providers have a very similar look and feel with regard to making complaints. In practice, TikTok's complaints interface should look very similar to Instagram or YouTube, for example.

It needs to be obvious and very clearly signposted so that this benefits all users, including children and users with a disability.

Question 17: What approach do you think the Code should take to ensuring that the safety measures we ask VSPS providers to take are accessible to people with disabilities?

See above.

5.3.2 Risk assessments

Where possible, the risk assessments required by the DSA and the OSMRA should be very aligned, consistent in approach and language and prescriptive (i.e. it should be not be up to the VSPS provider how they comply).

Risks to children should be a central consideration in any such risk assessment.

5.3.3 Safety by design

Safety by design should be a central design consideration for any platform that permits children on their platform. It would not be enough to simply ask VSPS providers to publish a statement setting out how they interpret safety by design on their service. This needs to be specified in a prescriptive way, within the Code. It must go further than a statement and it must be clear when this has been breached. We are not convinced that VSPS providers will truly adopt all necessary measures for a safety by design approached unless compelled to do so.

Question 18: What approach do you think the Code should take to risk assessments and safety by design? Are there any examples you can point us towards which you consider to be best practice?

See above comments on Risk Assessments.

Question 19: How do you think that cooperation with other regulators and bodies can help us to implement the Code for VSPS?

5.3.5 Harmful feeds and recommender systems

Bloggers and influencers on VSPS generate user content, often to the benefit of VSPS in terms of engagement, which can in some cases rely on the person's child promoting the page, promoting brands or activities. The use of children in such advertisements in Ireland is completely beyond the scope of regulation, because the child's guardian is posting the content. However, it is clear that there is a power imbalance often between the parent and the large brands who might be offering financial incentives to the parent. There should be cooperation between the OSC and the CCPC in



requiring that these brands act responsibly when entering into such agreements with parents for example, the child's privacy is often compromised. Also, the fair division of earnings from such branding is not provided for as is similarly in other jurisdictions. Rest breaks, prioritising the child's welfare and ensuring that the child's privacy can be maintained are often not considered. Given the power imbalance, there should be an obligation on brands via the CCPC to provide contractually for the child in terms of earnings. The expenditure by brands were such video content to be professionally made, employing child actors and promoted would far exceed what is being provided to smaller influencers, often at the expense of the child promoting the product. There should be guardrails in place as regards this type of marketing

Question 20: What approach do you think we should take in the Code to address feeds which cause harm because of the aggregate impact of the content they provide access to? Are there current practices which you consider to be best practice in this regard?

VSPS should be obliged to verify age of the users of their platforms and adjust their algorithmic feed accordingly.

We have covered harmful content and age-inappropriate content above however, it goes without saying that children should not be fed harmful content, such as food restricting content, pro-ana content, suicidal and self-harm content or sexual content. Children should not be fed illegal content and age-inappropriate content. Pornography is being algorithmically fed to young users on VSPS platforms. This is a fact that is not being in any way adequately addressed by platforms. The harms experienced by young users experiencing such content have been commented upon globally. This code is an opportunity to provide an obligation on VSPS to ensure their young users are identified on their platforms and not fed such harmful content.

As stated above, age-assurance measures should be used by VSPS providers to protect children on their services, this should mean no targeted or age-inappropriate advertising or content recommendations. This should be prioritised within the Code.

Regardless of whether or not the user is a child, there should be measures put in place to allow users to depersonalise their feeds, if they so choose. Such features should be clearly signposted on platform.

Question 21: Do you have any views on how requirements for commercial content arranged by a VSPS provider itself should be reflected in the Code?

5.3.7 Compliance

Effective age-verification requirements are key. Many VSPS will say that children under 13 are not permitted on their service. However, the CSK 2022 report (and many previous reports) speak to the fact that children under 13 are on these platforms in large numbers. It appears that marketing directed at young users can distinguish their ages from older service users. It also appears that age verification/assurance technologies are available. However, without mandatory requirements to do so, VSPS have no incentive to invest in such technologies. This is a key area of child protection



online. VSPS must be obliged to identify young platform users rather than simply deny their existence. This financial investment will not happen without mandates from the Regulator.

Annual compliance statement:

Debates prior to the OSMR Act showed that transparency by VSPS and indeed by many online services providers, was an issue. Annual reports are often vastly populated leaving very little by way of comprehension. Instead, perhaps annual compliance statements which include a set format with key areas to be addressed by each VSPS in a uniform manner to ensure each aspect of the Code requirements has been addressed in a coherent manner.

We note there are no criminal sanctions in the OSMR Act (as being proposed by the UK Online Safety Bill) and instead civil sanctions for non-compliance are outlined. In particular the OSMR Act provides for outline sanctions where the Codes are not adhered to. Where punitive measures are not legislated for, it would be important that the sanctions that apply for non-compliance with the Code are extensive. Commercial corporations listen to sanctions around reputation and financial consequences. We have seen this in the Irish DPC and global headlines for sanctions where laws were not adhered to. Similarly, the strength of the Code will lie in the penalties for non-compliance.

Question 22: What compliance monitoring and reporting arrangements should we include in the Code?

5.3.8 Transitional Arrangements

Transition periods For VSPS to comment.

Question 23: As soon as possible. More appropriate for VSPS comment

Prepared by:

- *Clare Daly, Solicitor with CKT, particular focus on child protection and data protection and CyberSafeKids Board member*
- *Alex Cooney, CyberSafeKids CEO*

Submitted on 4th September 2023



Dairy Industry Ireland
84/84 Lower Baggot Street
Dublin 2

T: +353 1 605 1500
E: info@ibec.ie
W: www.ibec.ie/dairyindustryireland

Coimisiún na Meán Online Safety Code call for inputs – Dairy Industry Ireland viewpoint

Dairy Industry Ireland (DII), the representative body for Irish primary and secondary dairy processors, including the infant nutrition sector, welcomes the opportunity to input to Coimisiún na Meán's first call towards the development of an Online Safety Code.

As a sector we strongly voice our support for the protection of children and young people from harmful online content, through codes and policy.

Commercial communications relating to infant and follow on formula have been referenced in the Online Safety and Media Regulation Act 2022 as a category of products for which commercial communication in audiovisual channels may be restricted or prohibited.

The regulation of such communications is already set down at Irish and EU level (including, but not limited to that referenced in footnote) and overseen by a range of national bodies. This includes laws governing written, verbal and electronic communication to consumers. Furthermore, DII member companies, which manufacture, and export these products have already shown commitment to voluntarily exceed compliance with such regulation, through own company codes and policies, as well as supporting the WHO's recommendation for exclusive breastfeeding in the first six months of life, followed by continued breastfeeding for up to two years and beyond.

The development of additional guidance, such as that developed by Dairy Industry Ireland jointly with the Food Safety Authority of Ireland on compliance with food law when communicating with health professionals about infant formula products ([link](#)), critically reflects the already stringent EU regulatory frameworks that already govern communication in relation to formula milks, and the willingness of the industry to engage with regulatory authorities and ensure strict compliance with the law relating to product communication

DII member companies all fully agree that breastfeeding is the best source of nutrition for babies and should be promoted and protected, with all necessary supports in place to do so. When breastfeeding is not possible or chosen, formula milks are the only legitimate and nutritionally complete alternative recognised by the World Health Organisation. When parents, caregivers or health professionals seek information on these products, it is essential that they are able to receive the most accurate and up-to-date guidance and advice.

DII member companies ask that any reference to infant and follow-on formula milks in the developed Code is evidence-based, proportionate and reflective of existing European regulation.

It is also our ask that discussion in relation to infant and follow-on formula milks would involve direct engagement with the industry and our members look forward to constructively providing science-based support and engaging collaboratively in this regard.

Regulation: 1) Commission Delegated Regulation (EU) 2016/127 regarding the specific compositional and information requirements for infant formula and follow-on formula and regarding the requirements on information relating to infant and young child feeding | 2) Regulation (EU) No. 1169/2011 on Food Information to Consumers | 3) Regulation (EC) no 1924/2006 of the European Parliament and of the Council on Nutrition and Health Claims made on Foods | 4) Regulation (EU) No 609/2013 of the European Parliament and the Council on food intended for infants and young children, food for special medical purposes, and total diet replacement for weight control.



Department of Health submission to Coimisiún na Meán Call for Inputs on *Online Safety – Developing Ireland’s First Binding Online Safety Code for Video-Sharing Platform Services.*

Table of Contents:

- Ministerial Foreword2
- Section 1: Mental Health.....5
- Section 2: Healthy Ireland.....13
- Appendix 1: Healthy Ireland Survey19



Ministerial Foreword

I welcome the opportunity to make a submission to Coimisiún na Meán to inform the development of Ireland's first binding Online Safety Code.

Recent years have seen the emergence of a growing body of international evidence identifying strong relationships between exposure to some online content and poorer mental and physical health outcomes amongst our young people in particular. The development of an Online Safety Code to address online harm arising from video-sharing platform services is a welcome step in tackling the dangers online content can pose to youth mental health.

At the outset it is important to recognise that present day modes of engagement with digital technologies are varied and continually evolving. Although widespread adoption of digital technologies has brought many benefits, there is a need to develop a more sophisticated awareness and understanding of the negative health impacts associated with excessive exposure to under-regulated, harmful and excessive online content.

The harmful psychological impacts of inappropriate exposure to online content, social media, and mobile phone use on youth mental health can include anxiety and stress, depression, self-harm, disordered eating, and suicidal ideation. In addition, we can see that excessive social media and mobile phone use and exposure to online content can have direct negative effects on interpersonal relationships as well as body image, through social comparison and from negative and harmful interactions online, such as cyberbullying. Similarly, heavy social media and mobile phone use can contribute indirectly to poorer health and wellbeing outcomes for young people, in particular, through sleep deprivation, and poorer academic and cognitive performance. High levels of screen use can also be associated with poorer physical health outcomes, and obesity, through long sedentary periods spent online. This array of potential negative health and wellbeing outcomes can have long lasting effects on psychological development, education outcomes, and long-term physical and emotional wellbeing.

In addition, it is widely documented that unhealthy food marketing- which is prevalent across all platforms- negatively affects taste preferences, food requests, food purchases, food consumption, and the nutritional quality of children's diets. The increased obesogenicity of these food environments has a consequently negative impact on health outcomes, including the risk of childhood obesity (WHO, 2021).



I am confident that the development of an Online Safety Code targeted at video-sharing platforms will help to address some of these issues. However, a holistic approach to safeguarding the health wellbeing of our youth will require further action across many forms of digital content.

Beyond online harm arising from content on video-platforms, we must also understand and address factors such as the spreading of misinformation through online channels, lack of personal age (and other) verification and accountability, limited transparency, and insufficient parental oversight and control over the types of online content entering their households.

Ultimately, the commercial incentive to maximise engagement with online platforms via the use of sophisticated marketing tools, such as artificial intelligence, is at odds with our responsibility to foster healthy behaviours and attitudes towards digital technologies amongst the youngest in our society.

An effective policy response to such a new and evolving health challenge will require whole of Government understanding and commitment. This should include improved monitoring of physical and mental health outcomes amongst our young people, as well as detailed examination of how our youth engages with digital technologies to inform the choice and design of policy interventions which are most suited to safeguarding and informing users of online content.

Work that is being undertaken by the Department of Health to support the role of the Commissioner, as well as legislative, regulatory, policy, and operational work that is already being undertaken include the following:

Summary of Submission by Mental Health Unit, Department of Health

With regard to online safety, the Department of Health leads on the development of online mental health tools and resources, working with the HSE and Healthy Ireland to signpost services and provide positive messaging about online activity.

Connecting for Life is Ireland's National Strategy to Reduce Suicide, and it aims to improve the nation's understanding of and attitudes to suicidal behaviour, mental



health, and wellbeing. The Strategy emphasises the importance of encouraging safer online environments and responsible reporting on suicide related content.

The Department believes the new code should address wider categories of harmful online content, such as content promotes or encourages self-harm or suicide or behaviour that characterises a feeding or eating disorder, and that harmful content related to suicide, self-harm and eating disorders in particular will need to attract the most stringent mitigation measures by VSPS providers.

Any classification or categorisation of harmful content should rate content related to suicide, self-harm and eating disorders as the most harmful, and should be specific in defining what types of content can fall within a particular category.

The Department of Health is supportive of the measures described in the EU Audiovisual Media Services Directive (including flagging mechanisms, age verification and increased parental controls) becoming part of the new online safety code. The Department believes that co-operation with other regulators and public bodies, including health services, will be essential to the implementation and effective operation of the Code.

Healthy Ireland- Obesity policy and the restriction of advertising of certain foods and beverages through media service codes and online safety codes

The policy instrument for obesity in Ireland is “A Healthy Weight for Ireland”, the Obesity Policy and Action Plan (OPAP), which was launched in September 2016 as part of the Healthy Ireland Framework. The establishment of An Coimisiún with a remit of developing media service codes and online safety codes represents a significant opportunity to drive the policy objectives of Healthy Ireland and the OPAP and to address in particular the prevalence of childhood obesity in Ireland.

The standards and practices that can be addressed through regulatory codes and rules developed by the Coimisiun na Mean include the advertisement of certain foods and beverages.

Ireland is currently working with European partners on a Joint Action under the EU’s Health Programme on a suite of supports setting out best practice with regard to developing binding codes or regulations, and monitoring and compliance in relation to



the restriction of advertising unhealthy or harmful foods and beverages to children. This work is due to conclude shortly.

Officials from Healthy Ireland have already engaged with the Coimisiún and will continue to work with the Coimisiún, in particular once the outcomes of the work at EU level are available to draw on. In the meantime, we are happy to provide a short chapter to the submission on the Call for Input: Online Safety.

Section 1 – Mental Health

With regard to online safety, the Department of Health leads on the development of online mental health tools and resources, working with the HSE and Healthy Ireland to signpost services and provide positive messaging about online activity.

Connecting for Life is Ireland's National Strategy to Reduce Suicide, and it aims to improve the nation's understanding of and attitudes to suicidal behaviour, mental health, and wellbeing. The Strategy emphasises the importance of encouraging safer online environments and responsible reporting on suicide related content. The National Office for Suicide Prevention (NOSP) within the HSE lead on implementation of *Connecting for Life*, and NOSP has also prepared a submission to this call for inputs from the Commission.

The Department and NOSP have been leading out on engagement with sectoral stakeholders including Samaritans, Headline, and the National Suicide Research Foundation (NSRF) to ensure they were aware of the call for inputs and to advocate that each organisation make its own submission to the call.

The Department also supports the implementation of the HSE National Clinical Programme for Eating Disorders (NCP-ED), a collaborative initiative between the HSE, the College of Psychiatrists of Ireland, and Bodywhys (the Eating Disorders Association of Ireland), the national support group for people with eating disorders. People with mental health problems, and notably people with eating disorders, have a heightened lifetime risk of, and vulnerability to, suicide. Suicide, self-harm and eating disorders are specifically referenced by the Broadcasting Act 2009 as potentially harmful content.

This submission responds to questions most aligned with the role and function of the DoH from the perspective of the Mental Health Unit.

3.1 What online harms should the Code address?



Question 1: What do you think our main priorities and objectives should be in the first binding Online Safety Code for VSPS? What are the main online harms you would like to see it address and why?

In addition to the harms addressed in Article 28b of the Audiovisual Media Services Directive, the Department of Health is of the view that the code should address wider categories of harmful online content, as detailed in the 2009 Broadcasting Act, including harmful online content on services by which a person:

- o Bullies or humiliates another person;
- o Promotes or encourages behaviour that characterises a feeding or eating disorder;
- o Promotes or encourages self-harm or suicide;
- o Makes available knowledge of methods of self-harm or suicide.

The 2009 Act as amended also specifies a further category of harmful online content relating to 42 criminal offences under Irish law listed in Schedule 3 of the 2009 Act as amended. Examples of offences include:

- o Non-consensual sharing of intimate images;
- o Child sex abuse material
- o Naming complainants in rape trials;
- o Material relating to suicide;
- o Harassment;
- o Child and human trafficking;
- o Domestic violence.

The promotion of suicide and self-harm is a key online harm which will need to be addressed through the development of a specific code. Any such code should address materials and information on different methods and rationales for suicide, any type of forum that encourages suicide, 'pact' websites, content (videos, images, descriptions) that depict suicide or self-harm acts.

Our understanding of the role social media can play in suicide clusters and increased ideation is increasing, and the code should specifically address this risk through requiring platforms to be proactive in identifying and removing harmful content.



Individuals with eating disorders have a heightened lifetime risk of suicide, and we know that body image concerns can be exacerbated by social media content and filtered/edited photographs. As such, the development of this code should address activities which promote or encourage behaviours that would characterise an eating disorder. Relevant harmful content would include pro-eating disorder websites which typically discuss, encourage or amplify concerning behaviours including how to conceal an eating disorder, resistance to treatment, weight loss strategies and challenges/competitions. Within the category of eating disorder related harmful content, the new Code should specify precisely what kinds of content this can be, such as the behaviours detailed above.

Question 2: What types of online harms do you think should attract the most stringent risk mitigation measures by VSPS? How could we evaluate the impact of different types of harms e.g. severity, speed at which harm may be caused? Is there a way of classifying harmful content that you consider it would be useful for us to use?

Harmful content related to suicide, self-harm and eating disorders in particular will need to attract the most stringent mitigation measures by VSPS.

In evaluating the impact of different types of harms, relevant issues would include considering if physical harm is likely to be caused, such as through self-harm, suicide, and also wider psychological impacts on individuals who may be traumatised or experience distress.

It must also be borne in mind that some groups may be more vulnerable to harm such as children, young people, those experiencing mental health difficulties, those who are suicide bereaved.

In terms of the speed at which harm could be caused, given the instantaneous nature of social media and messaging applications the Department believes that the code should assume that harm could be imminent/caused extremely quickly, and take account of that with regard to the timeframes afforded to platforms to address these issues. This point is also relevant to the issue of the amplification of content, such as through 'viral' videos which can increase the reach of content and become a public safety risk.

Any classification or categorisation of harmful content should rate content related to suicide, self-harm and eating disorders as the most harmful.

4.1 How prescriptive or flexible should the Code be?



Question 4: What approach do you think we should take to the level of detail in the Code? What role could non-binding guidance play in supplementing the Code?

The Department of Health would favour Option 3, the mixed approach. High level obligations setting out categories of harm and required mitigation measures, supplemented with more detail as appropriate may be the most effectively structured code.

Highly detailed or overly prescriptive regulatory frameworks can risk a ‘letter’ rather than ‘spirit’ approach from those being regulated, and encourage a narrower focus than a more flexible approach which requires platforms to reflect on key issues and consider how to reach compliance in a more proactive way and how they can demonstrate their own compliance.

Non-binding guidance would be essential in fostering this approach and attitude from platforms.

Transparent mechanisms such as the publication of data on the work platforms are undertaking to address harmful content and promote online safety could be effective.

Any code should also be responsive to emerging issues and be adapted to ensure that it remains effective in promoting online safety.

4.2 How should we structure the Code?

Question 5: What do you think would be the most effective structure for the Code? What are the most important factors we should consider when we decide how to structure the Code?

High level obligations setting out categories of harm and required mitigation measures, supplemented with more detail as appropriate may be the most effectively structured code.

A factor to consider in structuring the code would be the need to list categories of harmful content, but also within this to be specific in defining what types of content can fall within a particular category, for instance:

Suicide and self-harm content:

- Information on methods
- Pro-suicide and self-harm sites
- Online ‘games’



- Online imagery or videos of suicide and self harm
- Social media content which normalises self-harm and suicide, sharing of suicide notes, content about celebrity suicides which can increase risk.

As per our response to question 1, this need to define what types of content fall within a particular category will also apply to eating disorder related content.

Any code should also be structured in such a way that it can encompass harmful content which is not explicitly mentioned in the code, but which is judged to be harmful by the Commission as it occurs/on a case by case basis, and therefore requires action by the relevant platforms.

4.3 How should the Code take account of the Digital Services Act (“DSA”)?

Question 6: How should we design the Code to minimise the potential for conflict and maximise the potential for synergies in how platforms comply with it and the DSA?

The Code should take account of the DSA and if possible be structured in such a way that compliance with the Code equates also to compliance with the DSA / vice versa.

4.4 How should the Code address content connected to video content?

Question 7: To what extent, if at all, should the Code require VSPS providers to take measures to address content connected to video content?

To be truly effective, the Department believes that the Code should consider content connected to video content as potentially being as harmful as the video itself, and therefore requiring measures by VSPS. This is to reflect the fact that connected content, such as comments, could change the meaning or perception of video content, and make something more harmful.

5. Measures to be taken by Video-Sharing Platforms

The Department of Health would be supportive of the measures described in Article 28b.3 of the AVMSD that VSPS providers should take becoming part of the new online safety code.

Question 9: How should we ask VSPS providers to introduce and design a flagging mechanism in the Code? How can we ensure that VSPS providers introduce the mechanism in a user-friendly and transparent way? How should we ask VSP Providers to report the decisions they’ve made on content after it has been flagged?



To what extent should we align the Code with similar provisions on flagging in the DSA?

The Department is supportive of Commission plans to require VSPS providers to establish and operate transparent and user-friendly mechanisms for users to report or flag content in the Code, and to require VSPS providers to establish and operate systems to explain the decisions they make after content has been reviewed.

To encourage and support compliance, the Code should be as closely aligned as possible to the provisions on flagging within the DSA. VSPS platforms could be advised to integrate notification and flagging mechanisms.

With regard to harmful content related to suicide, self harm and eating disorders, information on appropriate supports and services should accompany flagging mechanisms for users.

Any information on suicide and self-harm supports should be responsive and relevant, i.e. it should be local to the person and time-specific (e.g. out of hours services may need to be signposted).

Question 10: What requirements should the Code include about age verification and age assurance? What sort of content should be shown by default to users who are logged out or in private browsing mode and whose age cannot be verified or assured? What evidence is there about the effectiveness of age estimation techniques? What current practices do you regard as best practice? Where accounts are not age verified should default privacy settings be used, should content default to universal content and should contact by others be more limited?

The Department would support the Commission's plan to require VSPS providers to introduce appropriate age-verification mechanisms to protect minors from online harms in the Code, and notes that Article 28b of the AVMSD requires content that is most harmful to minors to be subject to the strictest access control measures.

Age verification is less relevant with regard to very harmful content regarding suicide and self-harm, as this content should not be considered suitable for viewing by anybody and needs to be removed.

In terms of other content which could potentially cause harm, where a person's age cannot be verified the Code should err on the cautious side and only display universal content or content that is deemed suitable for the youngest users. In the absence of



verification, one must assume a young person or vulnerable person is viewing the content. This is to avoid the causing of harm in the first instance.

Question 14: How should we ask VSPS providers to address online harms in their terms and conditions in the Code, including the harms addressed under Article 28b? How should key aspects of terms and conditions be brought to users' attention? What examples are there of best practice in relation to terms and conditions including content moderation policies and guidelines?

VSPS provider terms and conditions should be considered as a tool to promote online safety. The Department would support the Code taking measures to prohibit harmful content related to self-harm, suicide and eating disorders. This should include a communication to users that this type of content is prohibited, and suitable sanctions for rule breakers such as account suspension/ termination. VSPS providers should be required to bring their terms and conditions to user attention in plain and easy to understand language. This is also an opportunity for providers to explain why such content is harmful.

Question 15: How should we ask VSPS providers to address content moderation in the Code? Are there any current practices which you consider to be best practice? How should we address automated content detection and moderation in the Code?

Content moderation can be a highly effective tool in protecting users from viewing harmful content related to suicide, self-harm and eating disorders, and therefore obligations should exist for VSPS providers to monitor such content.

The Department notes the reasons set out by the Commission as to why content moderation decisions can sometimes be inaccurate or contestable, however the Department believes that with regard to harmful content relating to suicide, self-harm and eating disorders, VSPS providers should be obliged remove this content, and guided to err on the side of removing content that relates to this area even if the instance seems less clear-cut.

The Department would favour VSPS providers being mandated to prioritise removal requests from certain bodies, such as other regulators, public bodies and health services. The Department is very supportive of HSE NOSP's suggestion around their direct engagement with VSPS providers to assist real-time detection and of responses to critical incidents or cases of suspected suicide.

The Department would favour high-risk content breaches resulting in rapid and appropriate sanction such as account suspension or termination.



Regarding harmful content related to suicide and self-harm in particular, the Department would favour specified timescales for VSPS provider decisions on flagged harmful content. Automated flagging should assist providers in adhering to these timescales. Timescales are also important as distress can occur when a platform does not swiftly act to review a notification by the user. There is a tangible risk of real-time harm occurring to more vulnerable users, requiring targeted obligations for the monitoring of such content.

Question 19: How do you think that cooperation with other regulators and bodies can help us to implement the Code for VSPS?

The Department believes that co-operation with other regulators and public bodies will be essential to the implementation and effective operation of the Code for VSPS. There are a range of public agencies and non-statutory bodies working in the area of suicide and self-harm prevention supports for people with eating disorders who VSPS providers should be required to cooperate with to improve understanding of appropriate responses, and on the alignment of codes with relevant public health information, such as information on suicide prevention and related supports.

Question 20: What approach do you think we should take in the Code to address feeds which cause harm because of the aggregate impact of the content they provide access to? Are there current practices which you consider to be best practice in this regard?

Feeds should be addressed in a similar manner to other online content in the Code. Taken individually, certain pieces of content may seem lower risk, however an individual may be viewing multiple such contents, and over a sustained period of time, in a feed, and so there is an aggregate risk and impact which needs to be addressed.

The Department believes that algorithms may be an effective tool to address this type of content, to minimise recurrence and links/further recommendations to harmful content. Reliable information on potential services and supports should feature in such feeds.

The Department would be supportive of VSPS providers being required to ensure their recommender systems do not result in feeds of content which in aggregate cause harm. Very 'negative' feeds (a risk for those who may be suffering from mental health difficulties or at risk of suicide) or feeds dominated by a certain type of content (e.g. fitness and beauty) should be intercepted by positive and supportive content, as part of the design of the platform.



Question 23: Should the Code have a transition period or transition periods for specific issues? Which areas touched on in this Call for Inputs may VSPS providers require time to transition the most? What time frame would be reasonable for a transition period?

The Department notes that the Commission anticipate including a transition period in the Code to give VSPS providers time to adapt to the new requirements. However, given the need for an Online Safety Code the Department would be supportive of having the shortest transition period necessary.

The Department further notes that transitional arrangements could apply to the entire Code or to specific provisions of it where appropriate; the Commission might consider whether parts of the Code dealing with the most harmful online content around suicide, self-harm and eating disorders should apply immediately.

Section 2 – Healthy Ireland

General comment:

In its European Region Obesity Report of June 2022, the World Health Organisation identifies restrictions on the advertisement of food and drink considered unhealthy or harmful to children in particular as one of the key policy tools to use in addressing the obesity epidemic.

The policy instrument for addressing obesity in Ireland is “A Healthy Weight for Ireland”, the Obesity Policy and Action Plan (OPAP), which was launched in September 2016 as part of the Healthy Ireland Framework. The OPAP covers a 10-year period up to 2025 and aims to reverse obesity trends, prevent health complications and reduce the overall burden for individuals, families, the health system, and the wider society and economy.

Ireland is currently working with European partners on a Joint Action called BestReMaP under the EU’s Health Programme on a suite of supports setting out best practice with regard to developing binding codes or regulations, and monitoring and compliance in relation to the restriction of advertising unhealthy or harmful foods and beverages to children. This work is due to conclude shortly.

The establishment of An Coimisiún, with its remit of developing media service codes and online safety codes, represents a significant opportunity to drive the policy



objectives of Healthy Ireland and the OPAP, in particular drawing on the findings by the Best ReMaP Joint Action.

The standards and practices that can be addressed through regulatory codes and rules developed by the Coimisiun na Mean include the advertisement of certain foods and beverages. In this regard, the OSMR Act states (in section 139k(5)) that codes and rules may prohibit or restrict the inclusion in programmes or user-generated content of commercial communications considered by An Coimisiún to be the subject of public concern in respect of the general public health interests of children, in particular infant formula, follow-on formula or those foods or beverages which contain fat, trans-fatty acids, salts or sugars.

The Act further that Coimisiún may consult with public health authorities in relation to proposed restrictions or prohibitions. Officials from Healthy Ireland have already engaged with the Coimisiún and will continue to work with the Coimisiún and other stakeholders as deemed appropriate, in particular once the outcomes of the work at EU level are available to draw on.

Set out below are inputs to a number of the questions posed in the Call document. At this early stage, it is not possible to provide detailed input in relation to the development, design or delivery of online safety codes with regard to the advertising of foods and beverages considered to be the subject of public concern in respect of the general public health interests of children. Such codes would also need to be developed in tandem with media service codes relating to this policy topic.

Question 1: What do you think our main priorities and objectives should be in the first binding Online Safety Code for VSPS? What are the main online harms you would like to see it address and why?

The marketing of unhealthy foods to children and adolescents is prevalent across all settings, and pervasive, despite the introduction of regulations and voluntary codes of conduct in most European countries, including Ireland (WHO, 2022). It is widely documented that unhealthy food marketing negatively affects taste preferences, food requests, food purchases, food consumption, and the nutritional quality of children's diets. The increased obesogenicity of these food environments has a consequently negative impact on health outcomes, including the risk of childhood obesity (WHO, 2021).



The HFSS foods that are predominantly marketed to children across all commercial communication channels include fast food products or take-away meals, sugar-sweetened beverages, chocolate, confectionary, salty and savoury snacks, sweet bakery items, breakfast cereals, dairy products, and desserts (WHO, 2022; DG SANTE, 2021).

On television channels, children aged 4-7 years are exposed in average to 4.7 spots/day for HFSS foods, drinks or quick service restaurants and children aged 13-17 years are exposed to 2.95 HFSS spots per hour. On digital media, children are exposed to HFSS marketing in social media, news media websites, and music and video streaming platforms, particularly the young people aged between 13 and 17 years (DG-SANTE, 2021). In adolescents, exposure to HFSS marketing is associated with a positive perception and norms regarding the consumption of such foods (WHO, 2022).

Section 139k(5) of the OSMR Act provides for the following: “Without prejudice to subsection (2) or (4), an online safety code may prohibit or restrict, in accordance with law, the inclusion in programmes or user-generated content of commercial communications relating to foods or beverages considered by the Commission to be the subject of public concern in respect of the general public health interests of children, in particular infant formula, follow-on formula or foods or beverages which contain fat, trans-fatty acids, salts or sugars.”

Prioritising the inclusion of such restrictions in the development of an online safety code would be a positive development in addressing childhood obesity and delivering on our commitments in the OPAP.

Question 2: What types of online harms do you think should attract the most stringent risk mitigation measures by VSPS? How could we evaluate the impact of different types of harms e.g. severity, speed at which harm may be caused? Is there a way of classifying harmful content that you consider it would be useful for us to use?

To be determined from outcomes from EU Joint Action on Best ReMap.

Question 3: Do you have reports, academic studies or other relevant independent research that would support your views? If you do, please share them with us with links to relevant reports, studies or research.

A significant amount of work has already been progressed at an EU level with regard to the marketing of unhealthy foods to children. Since 2020, Ireland has participated in a Work Package on Restricting the marketing of unhealthy foods to children and adolescents under the EU Joint Action “Best ReMaP” (Best practices in Reformulation, Marketing and public Procurement), working with 15 other Member States.



The EU Framework for Action is the final deliverable of the Joint Action Best-ReMaP Work Package 6, expected to be published at the end of September 2023. This Framework will contain all the tools developed by the Work Package, including best practices for restricting marketing of unhealthy foods to children. This work that has been carried out at an EU level and the outcome of same should assist in informing the development of codes relating to the advertising of food and beverages under the OSMR Act. It would be premature to finalise the details of the criteria for which foods and drinks, etc. might be included in any advertising restrictions in advance of the finalisation of this work at EU level. Healthy Ireland will engage with the Coimisiun following the publication of the work by BestReMaP.

Relevant documents and websites:

[Best-ReMaP – Healthy Food for a Healthy Future \(bestremap.eu\)](https://bestremap.eu)

[D6.2-Technical-guidance-for-codes-of-practice-to-reduce-unhealthy-food-marketing-to-children-in-EU-Member-States.pdf \(bestremap.eu\)](#)

[WHO European Regional Obesity Report 2022](#)

[gov.ie - Combatting Obesity \(www.gov.ie\)](https://www.gov.ie)

Questions 4-23

The remaining questions in the Call document are by and large quite technical and administrative and are listed below for information.

At this point in time, Healthy Ireland does not have any further input on questions 4-23 but would hope in due course to contribute to discussions on best practice for the development of codes relating to restriction of advertising once the work at an EU level is complete.

Question 4: What approach do you think we should take to the level of detail in the Code? What role could non-binding guidance play in supplementing the Code?

Question 5: What do you think would be the most effective structure for the Code? What are the most important factors we should consider when we decide how to structure the Code?

Question 6: How should we design the Code to minimise the potential for conflict and maximise the potential for synergies in how platforms comply with it and the DSA?



Question 7: To what extent, if at all, should the Code require VSPS providers to take measures to address content connected to video content?

Question 8: How should we ask VSPS providers to introduce a feature that allows users to declare when videos contain advertising or other type of commercial communications? Should the Code include specific requirements about the form in which the declaration should take? What current examples are there that you regard as best practice?

Question 9: How should we ask VSPS providers to introduce and design a flagging mechanism in the Code? How can we ensure that VSPS providers introduce the mechanism in a user-friendly and transparent way? How should we ask VSP Providers to report the decisions they've made on content after it has been flagged? To what extent should we align the Code with similar provisions on flagging in the DSA?

Question 10: What requirements should the Code include about age verification and age assurance? What sort of content should be shown by default to users who are logged out or in private browsing mode and whose age cannot be verified or assured? What evidence is there about the effectiveness of age estimation techniques? What current practices do you regard as best practice? Where accounts are not age verified should default privacy settings be used, should content default to universal content and should contact by others be more limited?

Question 11: What requirements should the Code have in relation to content rating? What do you consider to be current best practice? What experiences have you had using content rating systems on platforms and do you think they have been effective? What steps could we ask VSPS to take to ensure content is rated accurately by users?

Question 12: What requirements should the Code have in relation to parental control features? How can we ensure that VSPS providers introduce the mechanism in a user-friendly and transparent way? Can you point to any existing example of best practice in this area? Should parental controls be 'turned-on' by default for accounts of minors or where age is not verified?

Question 13: What requirements should the Code contain to ensure that VSPS provide for effective media literacy measures and tools?

Question 14: How should we ask VSPS providers to address online harms in their terms and conditions in the Code, including the harms addressed under Article 28b? How should key aspects of terms and conditions be brought to users' attention? What examples are there of best practice in relation to terms and conditions including content moderation policies and guidelines?



Question 15: How should we ask VSPS providers to address content moderation in the Code? Are there any current practices which you consider to be best practice? How should we address automated content detection and moderation in the Code?

Question 16: What requirements should the Code include about procedures for complaint-handling and resolution, including out-of-court redress or alternative-dispute resolution processes? To what extent should these requirements align with similar requirements in the DSA? What current practices could be regarded as best practice? How frequently should VSPS providers be obliged to report to the Commission on their complaint handling systems and what should those reports contain? Should there be a maximum time-period for VSPS providers to handle user complaints and if so, what should that period be?

Question 17: What approach do you think the Code should take to ensuring that the safety measures we ask VSPS providers to take are accessible to people with disabilities?

Question 18: What approach do you think the Code should take to risk assessments and safety by design? Are there any examples you can point us towards which you consider to be best practice?

Question 19: How do you think that cooperation with other regulators and bodies can help us to implement the Code for VSPS?

Question 20: What approach do you think we should take in the Code to address feeds which cause harm because of the aggregate impact of the content they provide access to? Are there current practices which you consider to be best practice in this regard?

Question 21: Do you have any views on how requirements for commercial content arranged by a VSPS provider itself should be reflected in the Code?

Question 22: What compliance monitoring and reporting arrangements should we include in the Code?

Question 23: Should the Code have a transition period or transition periods for specific issues? Which areas touched on in this Call for Inputs may VSPS providers require time to transition the most? What time frame would be reasonable for a transition period?



Appendix 1: Healthy Ireland Survey

Online Safety in Ireland: Evidence Base

Topic/Issue: Online Safety, Briefing Note on Evidence Base

Date: September, 2023

Unit: Health and Wellbeing Programme

Purpose: DoH Submissions to Coimisiún na Meán, briefing

Key Points:

- Young people increasingly active online, and are starting to use the Internet at younger ages. However, the OECD notes a lack of evidence definitively linking screen time to poor health outcomes in children and young people.
- Most screen time is defined as sedentary behaviour, with inactivity presenting significant health risks. The CSPPA study, 2022 finds that physical activity levels in children and young people have increased, with 23% of primary and 12% of post-primary students meeting the National Physical Activity Guidelines for children of 60 minutes of moderate to vigorous activity per day, 7 days per week (CSPPA 2018: 17% primary, 10% secondary).
- The most recent HBSC Report (2018) finds that, in Ireland, 8% of children and young people report cyberbullying; 13% of boys and 18% of girls report being cyberbullied.
- Recommendations on how families can reduce screen time include protecting sleep, prioritising face-to-face interaction and being aware of parents' media use, as children tend to learn by example. These and other factors are seen to be more important than taking a hard line over screen time limits to ensure the best start in life.



Screen Time

The 2022 Healthy Ireland Outcomes Framework report found that young people are increasingly active online, and are starting to use the Internet at younger ages. UK figures show that in 2020, nearly all children aged 5-15 and over eight in ten children aged 3-4 (82%) went online in 2020. Tablets were a key device for pre-schoolers: two thirds of 3–4-year-olds used them (67%), with around half owning one themselves (48%).¹

A recent ESRI / GUI survey on the experiences of children and young people in Ireland during the pandemic found that circa 50-65% of both 12-year-olds and 22-year-olds reported increases in screen time.² The most recent figures from the PISA study on screen time show that the percentage of students (age 15) using the Internet for more than six hours per day outside of school, during the school day, increased significantly between 2015 and 2018, rising from 13.6% to 20.1%. However, the OECD³ notes the lack of causal evidence linking screen time to negative child health, and that scientific research currently:

- is not conclusive enough to support evidence-based guidelines on optimal amounts of screen use or online activities and;
- does not provide evidence of a causal relationship between screen-based activities and mental health problems, although some associations between screen-based activities and anxiety or depression have been found.

Evidence-based guidelines from the UK⁴ pose the following four questions to be used by families to examine how they use screens. If families are satisfied with their responses, it is likely they are doing well regarding screen time: 1) Is screen time in your household controlled?; 2) Does screen use interfere with what your family wants to do?; 3) Does screen use interfere with sleep?; 4) Are you able to control snacking during screen time? Recommendations on how families can reduce screen time include protecting sleep, prioritising face-to-face interaction and being aware of parents' media use, as children tend to learn by example. These and many other

¹ Ofcom, Children and parents: media use and attitudes report 2020/21 (Ofcom, 2021), https://www.ofcom.org.uk/__data/assets/pdf_file/0025/217825/children-and-parents-media-use-and-attitudes-report-2020-21.pdf

² ESRI, Growing up in Ireland: special Covid-19 survey (GUI/ESRI, 2021): https://www.growingup.ie/pubs/Covid-KF_Web-ready.pdf

³ OECD, What do we know about children and technology? (OECD, 2019), <https://www.oecd.org/education/cei/Booklet-21st-century-children.pdf>

⁴ Royal College of Paediatrics and Child Health, The health impacts of screen time - a guide for clinicians and parents (RCPCH, 2019), <https://www.rcpch.ac.uk/resources/health-impacts-screen-time-guide-clinicians-parents>



factors are seen to be more important than taking a hard line over screen time limits to ensure the best start in life.

Cyberbullying

The 2018 Health Behaviour in School-aged Children study (HBSC) finds that 8% of children report ever taking part in cyberbullying. There are statistically significant differences by gender, age group and social class with boys, older children and children from lower social class groups are more likely to report cyberbullying others. Overall, boys (13%) are less likely than girls (18%) to report ever being cyberbullied, and younger children are less likely to report being cyberbullied than older children. Children from lower social class groups are more likely to report being cyberbullied than those from other social class groups.

Screen time, Physical Activity and Sedentary behaviour

The CSPPA Report 2022, published on 30th August, 2023, finds that physical activity levels in children and young people have increased since 2018, with 23% of primary school and 12% of post-primary students meeting the National Physical Activity Guidelines for children of 60 minutes of moderate to vigorous activity per day, 7 days per week (2018: 17% primary, 10% secondary).

Sedentary behaviour is defined as any waking behaviour with an energy expenditure of no more than 1.5 Metabolic Equivalent of Task (MET)⁵. Due to a lack of appropriate evidence internationally, definitive sedentary behaviour guidelines remain elusive. CSPPA data shows that boys are engaging more in video gaming than girls, whereas TV viewing is highest among post-primary girls, and the use of phones for social media doubles from primary to post-primary. Given the well documented links between sedentary behaviour and long-term health outcomes such as cardiovascular health, future research should capture the true extent and nature of sedentary behaviour in young people.

[REDACTED]

⁵ Children's Sport Participation and Physical Activity Study 2022: <https://www.sportireland.ie/sites/default/files/media/document/2023-08/CSPPA%202022%20Full%20Report.pdf>



Food Drink Ireland
84/86 Lower Baggot Street
Dublin 2
Ireland
D02 H720

T: +353 1 605 1500
E: info@ibec.ie
www.fooddrinkireland.ie

Coimisiún na Meán Online Safety Code call for inputs – Food Drink Ireland viewpoint

Food Drink Ireland (FDI), the main trade association representing the interests of over 150 food, drink and non-food grocery manufacturers and suppliers, welcomes the opportunity to input to Coimisiún na Meán's first call towards the development of an Online Safety Code.

As a sector we strongly voice our support for the protection of children and young people from harmful online content, through codes and policy. FDI member companies are committed to marketing and advertising their products responsibly and they operate rigorous internal marketing codes and initiatives in addition to complying with a comprehensive set of international, national and sectoral level codes and pledges. The industry's adherence to these advertising and marketing initiatives demonstrates its commitment to contributing to a healthier food environment.

Best-ReMaP is a Europe-wide Joint Action (2020-2023) that seeks to contribute to an improved quality of food supplied to citizens of Europe by facilitating the exchange and testing of good practices concerning policies relating to reformulation, labelling, marketing and procurement. The Department of Health is co-leading a Work Package around advertising and marketing of unhealthy foods aimed at children. While Best ReMap seeks to reduce children's exposure to the marketing of foods high in fat, salt and sugar (HFSS), including online, it does not call for a ban and is clear that self- and co-regulation, including through codes of conduct, should be used effectively.

The Audiovisual Media Services Directive (AVMSD), and its implementing legislation – the Online Safety and Media Regulation Bill – provides Coimisiún na Meán with a full suite of powers. Not every type of harm envisaged by the Directive requires the imposition of a binding online safety code. The Directive explicitly encourages the use of self- and co-regulation where appropriate – to meet the Directive's requirements. These regulatory approaches should be fully embraced by Coimisiún na Meán.

FDI member companies ask that any reference to foods or beverages containing fat, trans-fatty acids, salts or sugars in the developed code is evidence based, proportionate and based in research.

It is also our ask that discussion in relation to food and beverages would involve direct engagement with the industry and our members look forward to engaging constructively and collaboratively in this regard. We look forward to submitting more detailed inputs to future consultations.

ENDS

4 September 2023

FSM e.V.
Beuthstraße 6
10117 Berlin

T +49 (0) 30 240 484-30
F +49 (0) 30 240 484-59
office@fsm.de
fsm.de

Coimisiún na Meán

Call For Inputs: Online Safety Code

Introduction

The FSM is a German state-approved self-regulatory body for digital services and online media. Amongst our members are many of the Very Large Online Platforms as well as a range of video on demand services providers from across Europe.

Since 1997, the FSM has been working to ensure that children and young people can grow up with a safer and better Internet - in particular by combating illegal and harmful content online. To this end, the FSM operates an Internet Hotline that anyone can contact to report online content. In addition, the FSM does extensive educational work and promotes media literacy skills among children, young people and adults.

Having been selected as an observer to the Global Online Safety Regulators Network, we are dedicated to working together with regulators from around the world to help young people stay safe online while allowing for innovation and recognising the rapid development of our digital world.

We are thankful for the opportunity to give input for the development of Ireland's first binding Online Safety Code for video sharing platforms by Coimisiún na Meán.

We are aware that, while online harms are global by nature, the perspectives of young people and their parents might differ from country to country and that the results of research will not always be internationally consistent. Regulating providers whose services are available in different jurisdictions is therefore challenging. With this input, we draw from our experience as a Germany based organisation working together with global companies on a daily basis.

There are significant differences between the various video-sharing platform services (VSPS) available already today and probably even more so when looking at

Vereinsregisternr.: 20264 B,
AG Charlottenburg, Berlin
USt-IDNr. DE814341170

Bankverbindung:
Berliner Volksbank
BIC: BEVODE33
IBAN: DE51 1009 0000
7049 3160 08

services which will be developed in the future, regarding the number of users, their age, the size and focus of the platforms and the content that is being shared. Any regulation should therefore carefully balance mandatory requirements and optional measures. The Code should reflect this by taking a risk- and principle-based approach which is flexible enough so that different VSPS can employ the most appropriate instruments to protect and empower their users.

Question 1: Which harms to address

What do you think our main priorities and objectives should be in the first binding Online Safety Code for VSPS? What are the main online harms you would like to see it address and why? (Please remember that when we refer to ‘online harms’ and ‘online harm’ in this document this includes harm that can be caused by harmful online content, illegal content, inappropriate content and commercial communications collectively.)

We welcome that Coimisiún na Meán distinguishes between different qualities of online harms: Some content or behaviour will clearly be illegal while others might (only) be inappropriate for young people under a certain age. The obligations imposed by any regulatory measure should always reflect this in order to balance the fundamental rights of all citizens and the rights of children.

Recent studies such as our own [Youth Media Protection Index 2022](#) (“Jugendmedienschutzindex”) have shown the main online harms young people are worried about in general as well as individually confronted with when they use online services and platforms, including VSPS and social media platforms. Among the age group between 13 and 16 **the following online harms can be identified as most relevant**, especially because the number of young people being confronted with them has significantly increased compared to 2017 (first edition of FSM’s “Youth Media Protection Index” study).

- Being confronted with disturbing or scary content (48% of 13/14-year-olds and 63% of 15/16-year-olds have experienced this)
- Being the victim of cost traps, rip-offs or scams (27% of 13/14-year-olds and 42% of 15/16-year-olds have experienced this)
- Being incited to engage in risky behaviour (dangerous challenges, drug/alcohol use or self-harm (35% of 13/14-year-olds and 45% of 15/16-year-olds have experienced this)
- Being exposed to political or religious extremism (35% of 13/14-year-olds and 49% of 15/16-year-olds have experienced this)
- Meeting people online who cannot be trusted (46% of 13/14-year-olds and 60% of 15/16-year-olds have experienced this)

- Being bullied by others (51% of 13/14-year-olds and 53% of 15/16-year-olds have experienced this)
- Being harassed online (51% of 13/14-year-olds and 56% of 15/16-year-olds have experienced this)

This might be an indication of what issues young people would most likely want to see being addressed. This being said, some of these online harms will be more difficult to tackle from a regulatory perspective than others. For the purposes of the envisaged Code, a clear focus might best be put on risks which are mentioned in Article 28b of the AVMSD, thus avoiding conflicts with the scope of the DSA.

When considering whether to address CSAM (child sexual abuse material), regulators should take into account that from our practical experience there are hardly any reports about such content on VSPs. In most cases, CSAM will be automatically filtered out after being uploaded there, and other channels for spreading such content are much more relevant. However, the FSM Hotline does receive reports about content that is classified as CSEM (Child Sexual Exploitation Material). This includes children behaving in a sexually suggestive way in front of the camera, obviously under instructions they receive through chat or messenger services.

Although there is limited content on VSPs that can be classified as CSAM/CSEM and regulation is already strong, this topic should still be included in the code because of the severity of the offences.

Question 2: Classification of harmful content

What types of online harms do you think should attract the most stringent risk mitigation measures by VSPs? How could we evaluate the impact of different types of harms e.g. severity, speed at which harm may be caused? Is there a way of classifying harmful content that you consider it would be useful for us to use?

Most stringent risk mitigation should be applied in the following order:

- when there is actual ongoing harm, especially when content goes viral: abuse, life-threatening challenges, live-streaming of illegal acts
- criminal offences
- content not suitable for minors

Question 3: Studies and resources

Do you have reports, academic studies or other relevant independent research that would support your views? If you do, please share them with us with links to relevant reports, studies or research.

The [Youth Media Protection Index](#) („Jugendmedienschutzindex“) examines how the protection of children and young people from negative online experiences is reflected in the concerns, attitudes, skills and actions of parents as well as adolescents themselves.

As a result, strengths and weaknesses of the current media policy regulations for the protection of young people from harmful online media as well as the available media education support services become apparent. This empiric evidence offers a basis for further developments and optimisations. The study was initiated and published by the FSM and conducted by the Leibniz Institute for Media Research | Hans Bredow Institute (HBI) and the JFF – Institute for Media Research and Media Education in 2022.

The empirical basis of the Youth Media Protection Index is a representative survey of 805 children and young people in Germany aged 9 to 16 who use the internet. In each case, the parent who feels responsible for the children’s online use or online education was also interviewed. This study is a repeat survey. Empirical results were available for the first time in the form of the Youth Media Protection Index 2017. By using the same questionnaire for the most part, the data from both studies – from 2017 and 2022 – can be compared and constants as well as changes can be identified.

See presentation of study results:

https://www.fsm.de/files/2023/03/fsm_jmsindex_presentation_english-1.pdf

See complete study (German):

https://www.fsm.de/files/2023/01/fsm_jmsindex_2022_barrierefrei.pdf

Question 4: Prescriptiveness of the Code

*What approach do you think we should take to the level of detail in the Code?
What role could non-binding guidance play in supplementing the Code?*

While the Code needs to be prescriptive as such in order to be executable, it should also be flexible in order to accommodate a variety of different services and to encourage the best possible reaction by service providers. We have recently seen various research efforts by the platforms that led to different approaches to current challenges, and we appreciate that there are no one size fits

all solutions. Therefore, a mixed approach should be preferred. In addition, co- and self-regulatory measures should be encouraged, as foreseen in the AVMSD.

Question 5: Structure

What do you think would be the most effective structure for the Code? What are the most important factors we should consider when we decide how to structure the Code?

The Code could be structured along the Article 28b.3 measures of the AVMSD.

Question 6: Synergies with DSA requirements

How should we design the Code to minimise the potential for conflict and maximise the potential for synergies in how platforms comply with it and the DSA?

While focussing on the AVMSD, the Code would ideally be consistent and coherent with the requirements of the DSA.

The AVMSD includes provisions for content which is harmful or inappropriate for younger users but is not strictly illegal in a way that it would constitute a violation of criminal law, whereas the DSA, in its English language version, focusses on *illegal* content. It is left for the national regulators to determine what content they consider illegal in this regard. Other language versions, specifically the German, are less strict. Even though there is no precedent today, we expect the German understanding of “illegal content” under the DSA to extend to any types of content forbidden by law, even if only under certain conditions (e.g. content inappropriate for younger users which is not restricted by age assurance measures). From a user perspective, it will be difficult (yet not important) to determine on which legal grounds certain content is inadmissible. That is why when drafting this Online Safety Code, Coimisiún na Meán should have the upcoming execution of the DSA in mind.

Question 7: Comments and other content connected to videos

To what extent, if at all, should the Code require VSPS providers to take measures to address content connected to video content?

Many of the outlined online harms minors are confronted with occur in additional content or in the comment sections (see question 1). Especially comment sections tend to develop a momentum of their own. Even if a video itself is harmless, there is a possibility that the comments are not.

However, since the purpose of the Code will be the transposition of the AVMSD, mandatory measures should only be set for video content.

VSPS could be encouraged to apply optional safety measures for content which is connected to videos uploaded by users, though.

This is reflected already today in the way platforms allow their users to report illegal content or behaviour in videos or in the comments section alike.

Question 8: Declaration of advertisements

How should we ask VSPS providers to introduce a feature that allows users to declare when videos contain advertising or other type of commercial communications? Should the Code include specific requirements about the form in which the declaration should take? What current examples are there that you regard as best practice?

left unanswered.

Question 9: Flagging mechanisms

How should we ask VSPS providers to introduce and design a flagging mechanism in the Code? How can we ensure that VSPS providers introduce the mechanism in a user-friendly and transparent way? How should we ask VSP Providers to report the decisions they've made on content after it has been flagged? To what extent should we align the Code with similar provisions on flagging in the DSA?

There should be no difference between flagging mechanisms for AVMSD or DSA purposes. Again, users should not be required to choose from different methods based on different legal grounds.

It is important to inform users that they can report content or conduct they think is illegal. However, there will be more than one option for doing this in a user-friendly way, depending on the nature of the VSPS, the users' age and the way platforms are used. It therefore seems indeed advisable to demand user-friendly and transparent information but refrain from too strict provisions in the Code.

Users will want to know if their report was taken care of so the provider should always send an appropriate response, preferable not hidden in a support dashboard. Some services might want to send an email, others might find a different path. User feedback as well as VSPS's own research might be considered in order to find an appropriate balance between the expectations of the reporting persons and the feasibility of such solutions.

Since users tend to be disappointed if their report has not led to the removal of content flagged by them, platforms should always inform users of the reasons for their decision in a transparent and easily understandable way.

Question 10: Age verification and age assurance

What requirements should the Code include about age verification and age assurance? What sort of content should be shown by default to users who are logged out or in private browsing mode and whose age cannot be verified or assured? What evidence is there about the effectiveness of age estimation techniques? What current practices do you regard as best practice? Where accounts are not age verified should default privacy settings be used, should content default to universal content and should contact by others be more limited?

Recently we have seen an enormous development in the effectiveness of age estimation techniques. FSM member YOTI continue to be very vocal on their numbers (cf. <https://www.yoti.com/blog/yoti-age-estimation-white-paper/>). In December 2021, the FSM's independent expert commission thoroughly examined the age estimation system "Yoti Age Scan" and concluded that it meets the German legal requirements. This has been the first time the FSM saw fit to accept a tool for preventing the access to adult pornography by minors which did not require a personal identification and use of official documents, but merely relies on automatic age estimation. This might underscore the quality and feasibility of this fairly new approach.

Question 11: Content rating

What requirements should the Code have in relation to content rating? What do you consider to be current best practice? What experiences have you had using content rating systems on platforms and do you think they have been effective? What steps could we ask VSPS to take to ensure content is rated accurately by users?

Content rating systems can be an effective way to prevent minors from encountering inappropriate material online and, at the same time, enable all users to view content they would like to see.

It should be noted that often users themselves will not be able to provide precise age ratings like we know them from cinema, TV or VoD services. Asking users for a too granular rating is likely to lead to many wrong ratings. There might be services which target a diverse audience from all age groups. These services could encourage their users to label content which they think is not appropriate for all ages or a specific age group.

If VSPS providers are required to establish and operate easy-to-use systems that allow users to age-rate the videos they upload, it is important to ensure that VSPS providers take steps to help users understand content rating schemes.

It is also important to understand that the availability of age ratings might lead users (especially parents) to a sense of safety which is not necessarily consistent with the actual situation.

VSPS might want to offer their users options to flag ratings they think are incorrect, and a certain number of such flags might lead to a review by the service provider. Again, this will be different for each VSPS, so the Code should encourage such features yet not prescribe them in detail.

Question 12: Parental controls

What requirements should the Code have in relation to parental control features?

How can we ensure that VSPS providers introduce the mechanism in a user-friendly and transparent way? Can you point to any existing example of best practice in this area? Should parental controls be 'turned-on' by default for accounts of minors or where age is not verified?

What kind of parental control feature is appropriate and meets the needs of the users may vary greatly from platform to platform. Some might focus on screen-time whereas others may be used to restrict interaction with other users or limit the availability of certain content. It is therefore important to provide users with clear and transparent information on what tools are available and how they can be used.

A default-on setting is challenging: Such a setting will practically always require age assurance so that the service can be used in full. It seems favourable to encourage parents to make an informed decision and set up the parental controls the way they deem appropriate for their children. Age verification as a standard would most likely not be accepted by users.

Most of the VSPS available today do not specifically target adults and many explicitly exclude content which is inappropriate for minors. A default-on setting for these platforms would be overprotective.

Question 13: Media literacy

What requirements should the Code contain to ensure that VSPS provide for effective media literacy measures and tools?

The Code should include a general requirement that there must be media literacy measures for minors in place. These should include available comprehensible and transparent platform rules. Furthermore, VSPS should be asked to anticipate online harms and educate minors in a way that is appropriate for the target group (regarding content reception and production). VSPS should also explain available measures to strengthen media literacy (prevention and intervention) and their use. Available measures should be (easily) accessible. To increase visibility and actual use by minors, measures targeted at parents and educators should also be encouraged. The Code can provide concrete examples of implementation which are not binding.

Question 14: Terms and conditions

How should we ask VSPS providers to address online harms in their terms and conditions in the Code, including the harms addressed under Article 28b? How should key aspects of terms and conditions be brought to users' attention? What examples are there of best practice in relation to terms and conditions including content moderation policies and guidelines?

Terms and conditions should be phrased in a way that minors can easily understand them. If VSPS consider this challenging for legal reasons, they could provide minor-friendly versions of their terms and conditions labelled as supportive documents.

Question 15: Content moderation

How should we ask VSPS providers to address content moderation in the Code? Are there any current practices which you consider to be best practice? How should we address automated content detection and moderation in the Code?

The Code should clearly outline the expectations for content moderation, including the removal of illegal and harmful content. Given the purpose of the Code being the transposition of the AVMSD, measures should not interfere with requirements of the DSA.

VSPS providers should be encouraged to be transparent about their content moderation practices, including the use of automated systems, and provide regular reports on their efforts to combat harmful content. While automated content detection systems can be useful, they are not foolproof. VSPS providers should be

urged to have a robust human review process in place to ensure accurate and fair content moderation decisions.

It is important to note that best practices in content moderation are constantly evolving. Therefore, the Code should provide a framework that allows for flexibility and adaptation to new technologies and emerging challenges.

Question 16: Complaint handling mechanism

What requirements should the Code include about procedures for complaint-handling and resolution, including out-of-court redress or alternative-dispute resolution processes? To what extent should these requirements align with similar requirements in the DSA? What current practices could be regarded as best practice? How frequently should VSPS providers be obliged to report to the Commission on their complaint handling systems and what should those reports contain? Should there be a maximum time-period for VSPS providers to handle user complaints and if so, what should that period be?

Complaint handling and resolution requirements should be consistent with those in the DSA to ensure consistency and harmonization across regulatory frameworks.

From our work as a [self-regulatory body under the German NetzDG](#) we know that setting stringent timelines for complaint handling is challenging. While it is desirable from a user's perspective that platforms review complaints quickly, it is even more important that they make correct decisions in order not to limit the users' freedom of expression. If maximum time-periods are to be set, they should reflect that some infringements are easier to determine than others and, likewise, some infractions are more severe than others and therefore demand quicker reactions.

Question 17: Accessibility

What approach do you think the Code should take to ensuring that the safety measures we ask VSPS providers to take are accessible to people with disabilities?

People with disabilities should equally be considered when it comes to safety measures. Similar to how VSPS already provide some features to allow more accessible content, the provision of accessible and inclusive safety features for prevention and intervention as well as efforts to make them well-known amongst users should be encouraged.

Since Article 47 of the DSA encourages codes of conduct for accessibility at Union level, the Code should avoid any possible interference or discrepancies with these.

Question 18: Safety by design

What approach do you think the Code should take to risk assessments and safety by design? Are there any examples you can point us towards which you consider to be best practice?

A holistic safety by design concept is desirable. Incentives for this should be created. However, it seems difficult to make individual functions mandatory. Here, too, the Code must be able to adapt to the constantly changing technology and be flexible.

Providers should carry out a renewed risk assessment when they introduce new features on their platforms.

Question 19: (International) Cooperation

How do you think that cooperation with other regulators and bodies can help us to implement the Code for VSPS?

We are aware that, while online harms are global by nature, the perspectives of young people and their parents might differ from country to country and that the results of research will not always be internationally consistent. Regulating service providers that are available in different jurisdictions is therefore challenging.

We also know that VSPS struggle with making adjustments for only one market or country. Ideally, a constant and trustful dialogue between regulators leads to feasible solutions that work across Europe and even beyond. The Global Online Safety Regulators Network might be a good forum for such a dialogue.

Question 20: Feeds and recommender systems

What approach do you think we should take in the Code to address feeds which cause harm because of the aggregate impact of the content they provide access to? Are there current practices which you consider to be best practice in this regard?

left unanswered.

Question 21: Commercial content, advertisement (e.g. pre-roll)

Do you have any views on how requirements for commercial content arranged by a VSPS provider itself should be reflected in the Code?

left unanswered.

Question 22: Compliance

What compliance monitoring and reporting arrangements should we include in the Code?

left unanswered.

Question 23: Transitional arrangements

Should the Code have a transition period or transition periods for specific issues? Which areas touched on in this Call for Inputs may VSPS providers require time to transition the most? What time frame would be reasonable for a transition period?

left unanswered.

* * *



Coimisiún na Meán's public Call for Inputs on the development of Ireland's Online Safety Code for Video-Sharing Platform Services

Executive Summary

Thank you for the opportunity to respond to Coimisiún na Meán's (**CnaM**) public Call for Inputs on the development of Ireland's first binding Online Safety Code (**Code**) for Video-Sharing Platform Services (**VSPS**).

Google and YouTube have actively engaged with policymakers on the issue of regulatory oversight for content sharing platforms. We take user safety seriously and welcome the co-regulatory approach envisaged by the AVMS Directive. We have long been and remain committed to tackling illegal and harmful content online and were pleased to participate in the Irish Government's 2019 consultation on the regulation of harmful online content, which preceded the Online Safety and Media Regulation Act (**OSMR**).

Deciding what content is allowed on YouTube, whilst avoiding undue interference with the fundamental right to free expression – including the right to seek, receive and impart information online – is a significant responsibility. We have implemented measures in line with Article 28b of the AVMS Directive, such as systems for reporting and removing harmful content and providing user tools to mitigate risks. These initiatives were developed with a multi-disciplinary team, in consultation with civil society, academia, and regulators.

In our submission, we advocate for a principles-based and non-prescriptive Code, in line with the AVMS Directive. This approach would offer flexibility and encourage continued innovation while providing robust user protection. It would also ensure a level playing field across the EU Digital Single Market and be designed to be future-proof.

In our response, we outline our views that:

- The main priorities and objectives in the first binding Code for VSPS should be the effective and swift transposition of Article 28b of AVMS Directive, as an urgent priority for Ireland.
- The first Code should focus on a clear set of rules for what constitutes illegal content; standards for transparency and best practices; oversight of systemic failures; and international cooperation.
- Section 3.1 of the Call for Input rightly identifies the four key areas of concern in the AVMS Directive that should be addressed by the Code: the protection of minors from content which may impair their physical, mental or moral development; the protection of the general public from incitement to violence or hatred content; content that contravenes EU law (such as public provocation to commit a terrorist offence); and, certain commercial communications that would not be permitted on broadcast or video-on-demand services.

- The Code should not go beyond these four key areas, in light of the specific focus of the Code on the provisions of the AVMS Directive and, to avoid fragmentation of the Digital Single Market and overly burdensome requirements.
- The Code should enable VSPS to enact proportionate risk mitigation measures.
- The Code should reflect and be consistent with the laws governing online services in the EU, in particular the Digital Services Act (**DSA**) (where YouTube has been designated as a “very large online platform”), the E-Commerce Directive and across other relevant frameworks in other fields, such as privacy and data protection, existing EU level codes of conduct, and jurisdictions and applicable law.
- The Code’s structure should also reflect principles-based regulation, focusing on the requirements of Article 28b of the AVMS Directive, with core sections addressing content categories and appropriate measures supplemented by non-binding guidance for adaptability and effective regulation.
- Extending AVMS Directive measures to ancillary features like comments would be disproportionate and go beyond its intended scope. Oversight for issues concerning such ancillary features should fall under the DSA’s risk assessment regime to maintain a harmonised approach and avoid conflicting regulation. We are also concerned about any further delays in fully implementing the AVMS Directive that might arise were the Code to introduce requirements that go beyond those envisaged under Article 28b.
- Platforms are best suited to design appropriate user-friendly, transparent disclosure tools for commercial communications specific to their individual services. While the Code should outline core principles for transparency and standardisation, it should also allow for platform-specific adaptations, recognising that different VSPS have unique functionalities and responsibilities to users.
- The Code should reflect existing age verification guidance for the protection of minors online as set out in the Irish Data Protection Commission’s *“Fundamentals for a Child Orientated Approach to Data Protection”* to ensure measures are appropriate and proportionate.
- Responding to harmful content is a dynamic challenge that requires constant adaptation. The Code should avoid being prescriptive about specific content moderation practices, with non-binding guidance that encourages innovation and remains flexible as such practices continue to develop and in response to new and emerging risks.
- While there are potential challenges that may arise from multiple regulatory input, cooperation from different stakeholders can enrich Code implementation. A principles-based Code that mirrors AVMS Directive requirements can ensure consistency across the Digital Single Market.
- A regulatory monitoring framework in the Code should focus on structural compliance rather than prescriptive reporting requirements which would allow for a proportionate and more effective approach.
- Timeframes for implementation should be sensitive to the final shape of the Code. Effective and sensible transition periods are crucial for VSPS and we recommend that CnaM consider a timeframe which will allow providers to get implementation right upon being well-informed of the forthcoming obligations.

The DSA expressly warns Member States against adopting additional national laws on the matters covered by the DSA, given that “*diverging national laws negatively affect the internal market*”, and emphasises the importance of the uniform application of its harmonised rules, so as to “*put an end to fragmentation of the internal market*” and “*ensure legal certainty*”¹. In particular:

- CnaM should ensure that the Code aligns with the risk assessment provisions in the DSA to avoid duplication and the possibility of divergent rules across jurisdictions seeking to achieve similar overarching objectives.
- Extending AVMS Directive measures to ancillary features like comments would be disproportionate and go beyond its intended scope. Oversight for issues concerning such ancillary features should fall under the DSA’s risk assessment regime to maintain a harmonised approach and avoid conflicting regulation. We are also concerned about any further delays in fully implementing the AVMS Directive that might arise were the Code to introduce requirements that go beyond those envisaged under Article 28b.
- Redress and complaint handling procedures in the Code should align with requirements in the DSA to minimise friction and avoid unnecessary duplication. CnaM should therefore avoid establishing a parallel, competing process for users which would result in legal uncertainty and confusion. In particular, any out-of-court redress for complaints about individual content moderation decisions should fall within the remit of the DSA.
- Article 47 of the DSA promotes the creation of codes of conduct by the European Commission to enhance accessibility for individuals with disabilities. The Code should adopt a flexible, non-prescriptive approach to allow easy adaptation in line with these future codes of conduct.
- CnaM should also take into account the transparency obligations under the DSA when designing and implementing transparency obligations under the Code. Given the scale of content made available on VSPS, regulation should focus on systemic failures when decisions regarding content are made. The Code should consider procedural accountability where evidence indicates systemic failures.

Open platforms such as YouTube are a benefit to a thriving audiovisual sector, providing a platform for European creators to find new audiences and providing viewers with access to information, supporting the free flow of ideas. YouTube is an important driver for creativity, enabling people to share their talents and raise their voices, across and outside Europe. European partners of all kinds and sizes are succeeding on YouTube, including new European talent, established creative industries, cultural institutions and other creative entrepreneurs. These partners are using the platform to communicate, entertain, educate, promote tolerance and understanding, and make a living. An overly prescriptive and non-harmonised approach to the Code puts this at risk.

We look forward to working with you in the further development of a Code that will ensure the best safeguards for our users.

¹ Recital 2, 4 and 9 DSA

RESPONSE TO CONSULTATION QUESTIONS

Section 3 – Online Harms

Question 1: What do you think our main priorities and objectives should be in the first binding Online Safety Code for VSPS? What are the main online harms you would like to see it address and why?

Section 3.1 of the Call for Inputs rightly identifies the four key areas of concern in the AVMS Directive that should be addressed by the Code:

- The protection of minors from content which may impair their physical, mental or moral development;
- The protection of the general public from content containing incitement to violence or hatred,
- The protection of the general public from content which is a contravention under EU law, such as public provocation to commit a terrorist offence; and
- The protection of the general public from certain commercial communications that would not be permitted on broadcast or video-on-demand services.

These categories of content are the specific focus of Section 139K(3) of the Online Safety and Media Regulation Act 2022 which relates to VSPS.

YouTube takes these categories of content extremely seriously and already has long standing policies and procedures in place to tackle such content expeditiously. Users of our platforms must follow rules of conduct, and we provide mechanisms for users to report such content.

A. What do you think our main priorities and objectives should be in the first binding Online Safety Code for VSPS?

In our view, the main priorities and objectives in the first binding Code for VSPS should be the effective transposition of Article 28b of AVMS Directive, as the transposition of this legislation is an urgent priority for Ireland.

We urge CnaM to ensure the Code remains focused on implementing the AVMS Directive and mirrors its approach of imposing obligations on VSPS without applying prescriptive and potentially overly burdensome requirements on platforms, as would that risk further delays and uncertainty with Ireland's transposition of these measures.

In particular, the AVMS Directive focuses on platforms taking appropriate measures to: (a) protect all users from 3 types of content: terrorist content, hate speech and child sexual abuse material (**CSAM**); and (b) protecting minors from content that may "*impair the[ir] physical, mental or moral development*".

We believe that, taking into account relevant differences between services, CnaM's Code should be guided by four principles:

- it should clearly set out the rules on what constitutes illegal content;
- it should set out standards for transparency and best practices;
- it should address systemic failures of Code provisions when responding to identified violations, not specific individual failures; and
- it should foster international cooperation, while recognising appropriate national differences.

To achieve this, we believe that CnaM should implement a principles-based Code that is proportionate, risk based and supports the role of co-regulation. We note that this approach has successfully been followed by other EU Member States, for example KommAustria requires "*platform*

providers ... to set up their own systems to ensure compliance with regulatory objectives (such as preventing online hate speech or child pornography). In line with the principle of proportionality, the regulatory authority only intervenes if a systemic/structural failure of a system set up by a platform is identified ("co-regulation" system)".

B. What are the main online harms you would like to see it address and why?

When it comes to both illegal and legal but harmful content, we urge CnaM to align the Code with the requirements of the DSA to avoid unnecessary duplication or overlap in regulating online content and to ensure consistency across the Digital Single Market (see answers 6, 9 and 18 in particular).

In addition, if the AVMS provisions are to take EU-wide effect, the inclusion of content defined by reference to Irish criminal law offences could undermine the harmonised approach required by this regime.

Question 2: What types of online harms do you think should attract the most stringent risk mitigation measures by VSPS? How could we evaluate the impact of different types of harms e.g. severity, speed at which harm may be caused? Is there a way of classifying harmful content that you consider it would be useful for us to use?

We believe the Code should focus on systemic matters in relation to lawful but harmful content, rather than seeking bespoke arrangements to address individual harms.

The DSA makes an important distinction between illegal content and "legal but harmful" content and the explanatory memorandum to the DSA Proposal recognised that the removal of "*harmful' (yet not, or at least not necessarily, illegal) content ... is a delicate area with severe implications for the protection of freedom of expression*". We would expect the Code to recognise the need to give due consideration to freedom of expression rights (noting the unique characteristics of user-generated content) and we would encourage an explicit emphasis on protecting the benefits that innovation in the VSPS market can bring to users.

The Code should respect this clear policy intention of the DSA to protect the rights of EU citizens to engage in debate and access lawful information, even in relation to issues that are contentious, and to ensure that appropriate checks and balances are in place in relation to any measures that impinge upon freedom of expression and access to information rights. We therefore consider it appropriate for the Code to apply less prescriptive requirements on these categories of content, enabling VSPS to enact proportionate risk mitigation measures - YouTube has existing policies, processes and systems in place to ensure that such content is prohibited and is expeditiously removed where it is deemed necessary, or where content is potentially harmful to children, it is age-restricted. The DSA recognises this and achieves an appropriate balance by adopting a systematic approach to the identification and mitigation of relevant risks.

In addition, we urge CnaM to ensure consistency with other codes to avoid undermining the policy intention behind the DSA's harmonised rules. If the Code is inconsistent with codes of conduct drawn up under the DSA and existing EU level codes, such as the Hate Speech Code and the Code on Disinformation that are well established and led at EU level, it could lead to a fragmented approach, with legal uncertainty.

Question 3: Do you have reports, academic studies or other relevant independent research that would support your views? If you do, please share them with us with links to relevant reports, studies or research.

YouTube launched its Community Guidelines Enforcement Report in 2018 to increase transparency and accountability around our efforts to keep users safe and these reports may assist CnaM in the development of the Code (see https://transparencyreport.google.com/?hl=en_IE). We have continued to publicly share a range of additional metrics for YouTube, such as the [Violative View Rate](#), to give more context about our work to protect users from harmful content.

Section 4 - Overall Approach to the Online Safety Code

Question 4: What approach do you think we should take to the level of detail in the Code? What role could non-binding guidance play in supplementing the Code?

We note that good regulatory practice, as identified in Ireland's Better Regulation policy², should leave "maximum flexibility as to how goals can be achieved". This highlights the importance of setting clear goals in the Code, but enabling VSPS to deliver upon those goals based on the specific characteristics and functionalities of their service.

To achieve this, we believe that CnaM should implement a Code that is proportionate, risk based and focused on the principle of co-regulation.

Given the variety of VSPS that exist, and which the Code will regulate, detailed prescriptive rules on the terms of service and the design of products and features would in our view:

- Risk fragmenting the Digital Single Market with VSPS under the jurisdiction of other EU Member States being subject to a materially different approach;
- Fail to recognise or support the risk based approach applied to larger platforms under the DSA;
- Fail to ensure the proportionate approach enshrined in the AVMS Directive which recognises that the regulation of VSPS should reflect "the size of the video-sharing platform service and the nature of the service that it provides"³;
- Hinder the ability of the rules to grow and adapt with EU law developments such as the DSA codes of conduct;
- Potentially harm innovation to products, limiting the possible evolution of technologies and helpful safety features;
- Fail to effectively regulate the broad range of services that meet the VSPS definition; and
- Increase the risk of conflict arising between the requirements of the Code and existing EU regulatory frameworks that apply to VSPS, including the DSA.

We note that a non-prescriptive approach to the implementation of AVMS Directive reflects the approach adopted in many other Member States⁴, and any deviation from this approach should be carefully considered. It is also worth noting the report recently published by the VSPS regulator in the UK, Ofcom, on the terms and conditions of VSPS registered in the UK, where it is stated that "highly detailed explanations of how terms and conditions are implemented may create opportunities for users to circumvent the rules and post harmful content".⁵

² <https://www.gov.ie/pdf/?file=https://assets.gov.ie/3477/281118144439-cf60aac3e3504e6f9f62f0ccda38f203.pdf#page=null>

³ Article 28b(3) of the AVMS Directive

⁴ "Mapping of national rules applicable to video-sharing platforms" the European Audiovisual Observatory

⁵ https://www.ofcom.org.uk/data/assets/pdf_file/0025/266173/VSP-user-policies-report.pdf

We therefore believe that the Code should be principles based and focused primarily on the core requirements of the AVMS Directive, as envisaged by Section 139K(3) of the Online Safety and Media Regulation Act 2022, with detailed implementation managed by VSPS themselves specific to the risks posed on each individual platform. Non-binding guidance could then be used to share more prescriptive best practice on these principles.

Question 5: What do you think would be the most effective structure for the Code? What are the most important factors we should consider when we decide how to structure the Code?

As noted above, we believe that CnaM should implement a Code that focuses on the requirements of Article 28b of the AVMS Directive without applying prescriptive and potentially burdensome requirements on VSPS. This will enable the development of a proportionate, risk based Code which is based on the principle of co-regulation.

For the reasons outlined in our response to Question 4, the suggestion in the Call for Inputs of a “very high level Code” that “may have only one or two sections” seems the most appropriate. This could mirror the approach of Article 28b of the AVMS Directive which focuses on (i) the categories of content VSPS must address and (ii) the appropriate measures they may take in relation to that content.

A further section of the Code could relate to compliance, information sharing and dispute settlement. The Code could then be supplemented by non-binding guidance where required, which would allow for greater adaptability and proportionality, and accordingly more effective, future-proofed regulation of VSPS.

Question 6: How should we design the Code to minimise the potential for conflict and maximise the potential for synergies in how platforms comply with it and the DSA?

The Call for Inputs provides that the Code might “...impose additional and/or more detailed requirements on VSPS providers” over and above the requirements under the DSA.

Where this goes beyond measures required by the AVMS Directive, this approach risks cutting across the harmonised approach required by the DSA.

More particularly, the DSA specifically recognises that the provisions of the AVMS Directive regarding VSPS should apply alongside the DSA to the extent that they regulate “specific aspects of provision of intermediary services”.⁶

To ensure the effective implementation of the AVMS Directive and to avoid duplication of regulatory requirements, confusing and divergent processes for users, and additional burdens and costs on businesses, it is crucial that the Code does not conflict in any manner with the DSA, including by the imposition of additional and/or more prescriptive requirements than permitted by the DSA.

We support the points made in the Call for Inputs that CnaM would seek “to minimise the potential for conflict and maximise the potential for synergies in how platforms comply with it and the DSA”.

In our view, the clearest way to achieve this is by ensuring the Code is principles-based, reflecting the high level requirements of the AVMS Directive. This would ensure that the Code encourages the adoption of appropriate measures by VSPS in areas where the AVMS Directive and the DSA regulate in parallel, and ensures that compliance with both regulations can be achieved in a harmonised way. More detailed provisions could also be included within non-binding guidance. This will support a harmonised approach across the Digital Single Market, as the DSA seeks to achieve.

⁶ Recital 10 DSA

This would also ensure consistency with the AVMS framework and the DSA as it continues to be implemented, particularly since the DSA provides for the drawing up of voluntary codes of conduct at EU level which will facilitate the proper and consistent Union-wide application of the DSA. Non-binding guidance can be updated more easily (where required) to ensure consistency against such codes as they emerge.

We note that, in its previous submissions to the Joint Committee on Enterprise, Trade and Employment and to the Irish Government in relation to the OSMR, Technology Ireland highlighted particular concerns about the conflict and overlap between the AVMS Directive, OSMR and the DSA in relation to the following areas:

- Content limitation orders: 'Legal but harmful' content
- Trusted flaggers/nominated bodies
- Complaints-handling and dispute settlement
- Transparency reporting
- Risk assessments and audits
- Blocking orders
- Fines and penalties
- Criminal sanctions as an enforcement mechanism

There are many examples where conflicts between the DSA and the Code may emerge. For instance, the DSA only allows Member State authorities to order the removal of illegal content. While the DSA includes some requirements in respect of lawful content, such as an obligation on the largest services to assess and mitigate systemic risks to users, including from certain legal but harmful content such as disinformation, it makes clear that *"harmful" (yet not, or at least not necessarily, illegal) content ... should not be subject to removal obligations, as this is a delicate area with severe implications for the protection of freedom of expression.*" (see the DSA's Explanatory Memorandum)⁷.

Notwithstanding the above, if any actions were required to be taken in respect of "legal but harmful" content pursuant to a Content Limitation Notice, service providers would need to understand how the DSA user redress regime (Articles 17, 20 and 21) - in respect of "restrictions imposed" and "decisions taken by" the provider of an online platform - would apply in respect of such actions. The potential for Content Limitation Notices in respect of "legal but harmful" and illegal content to apply pursuant to obligations set out under the Codes may both overlap with the regime set out in the DSA and undermine the harmonised approach to such issues that the DSA is intended to achieve. A prescriptive approach adopted under the Codes would be likely to heighten these concerns, more particularly where it brings into scope categories of harm outside of those envisaged by AVMS.

Question 7: To what extent, if at all, should the Code require VSPS providers to take measures to address content connected to video content?

The AVMS Directive specifies that the "appropriate measures" VSPS should take to protect users apply to *"programmes, user-generated videos and audiovisual commercial communications"*⁸, as opposed to purely ancillary features, such as comments.

We would note that YouTube takes issues with content connected to video content seriously. For example, between January and March 2023, YouTube blocked over 850 million comments, detected through a mix of automated and human flagging.

⁷ <https://eur-lex.europa.eu/legal-content/EN/TEXT/HTML/?uri=CELEX%3A52020PC0825>

⁸ AVMS Directive Article 28b(1)(a)-(c)

In our experience, comments and connected ancillary content generated by other users are typically viewed to a much lesser degree than video content, and therefore pose a lower risk of exposure to the general public and a lower risk of general harm. In the EU, users spend less than 1% of their time on YouTube engaging with comment functionality (as of Q4 2022). Nonetheless, YouTube's existing policies and processes - including reporting tools and removals - extend to comments and other features connected to a video, such as the thumbnail or a link in the video description. YouTube also offers creators the ability to turn off or to moderate comments on their videos.

We believe that extending any out-of-court redress mechanisms to individual user complaints about such ancillary features would be disproportionate, place an unnecessary burden on platforms, and would extend beyond the intended remit of the AVMS Directive. Again, this risks cutting across the harmonised approach required by the DSA.

Section 5 Measures to be taken by Video-sharing Platforms

5.1: Online Safety features for users

Question 8: How should we ask VSPS providers to introduce a feature that allows users to declare when videos contain advertising or other types of commercial communications? Should the Code include specific requirements about the form in which the declaration should take? What current examples are there that you regard as best practice?

VSPS should be encouraged to adopt a practical approach when introducing features for declaring commercial communications over which the VSPS does not have control and are not therefore "marketed, sold or arranged" by the VSPS. This would also be in line with the AVMS Directive which states that, in requiring VSPS to take these measures, "the limited control exercised by those video-sharing platforms over those audiovisual commercial communications" should be taken into account. We suggest that the rules make it clear that platforms are best placed to design their own appropriate product features, provided that: (i) the tools are easy for creators to use; and (ii) the disclosure is clear, up-front and sufficiently prominent to users.

At YouTube, we provide clear policies for creators featuring paid product placements, sponsorships, and endorsements ("**paid promotions**"). These measures are set up to offer certainty, clarity and a consistent experience for all stakeholders, including uploaders, viewers, and advertisers.

Through our experience in hosting commercial communications and content featuring paid promotions, we have developed policies that we consider to be best practices. These practices include:

- Clear policies regarding paid product promotions, sponsorships, and endorsements.
- Automatic disclosure messages overlaid at the beginning of videos.
- Encouraging uploaders to understand the laws and regulations around paid promotion in their jurisdictions.

While there should be flexibility in how VSPS disclose declarations for paid promotions, we agree that the Code should outline principles to ensure transparency for uploaders and viewers as well as standardisation principles that still takes account of the diverse VSPS landscape. On YouTube, we require creators to select the 'paid promotion' box when uploading videos containing commercial communications, which triggers an automatic disclosure message for 10 seconds at the start of such videos. However, this is just a baseline. We also remind creators that they must comply with any local advertising rules that may require additional disclosures. This approach provides a clear and

consistent experience for users, but still recognises that creators may be subject to a range of obligations from ads regulators in their own territory.

Whilst these practices work on YouTube, other platforms have developed ways unique to their own functionality. Whatever CnaM judges to be best practice should also reflect that each VSPS will engage with commercial communications in a unique way and allow the provider the flexibility to collectively ensure a safe, trustworthy and transparent experience for users.

Question 9: How should we ask VSPS providers to introduce and design a flagging mechanism in the Code? How can we ensure that VSPS providers introduce the mechanism in a user-friendly and transparent way? How should we ask VSP Providers to report the decisions they've made on content after it has been flagged? To what extent should we align the Code with similar provisions on flagging in the DSA?

As acknowledged in the Call for Inputs, what works best as a flagging mechanism may vary from service to service.

YouTube relies on a combination of people and technology to flag inappropriate content. Flags can come from our automated flagging systems, from members of our Priority Flagger program, or from users in the broader YouTube community. Additionally, YouTube is also required to comply with the trusted flagger requirements under the DSA.

The Code should align with similar provisions on flagging in the DSA⁹. Any notice and action mechanisms for users to notify a VSPS of a piece of content which it believes to be illegal content should align with the requirements of Article 16 of DSA.

We recognise that flagging tools should be easy to find and easy to understand for users. The decision making process should also be transparent, so that users are provided with clear information about removal processes, how to submit complaints, and how to assess and act upon those submissions.

Given the scale of content made available on VSPS, and the fact that individual user redress is addressed extensively under the DSA, we are of the view that regulators should not focus on decisions about individual pieces of content, but rather should consider systemic failures when decisions regarding content are made. The focus of the Code should therefore be on ensuring procedural accountability and regulating matters where evidence indicates systemic failure. Transparency will be paramount here, and we would encourage CnaM to take into account the transparency obligations under the DSA before introducing any transparency requirements under the Code. This will ensure that there is no conflict or overlap between the requirements of the Code and the provisions of the DSA.

Question 10: What requirements should the Code include about age verification and age assurance? What sort of content should be shown by default to users who are logged out or in private browsing mode and whose age cannot be verified or assured? What evidence is there about the effectiveness of age estimation techniques? What current practices do you regard as best practice? Where accounts are not age verified should default privacy settings be used, should content default to universal content and should contact by others be more limited?

⁹ Recital 10 DSA: 'However, to the extent that those Union legal acts pursue the same objectives as those laid down in this Regulation, the rules of this Regulation [i.e. the DSA] should apply in respect of issues that are not addressed or not fully addressed by those other legal acts as well as issues on which those other legal acts leave Member States the possibility of adopting certain measures at national level.'

Age verification and age assurance requirements in the Code should reflect wider, ongoing policy developments in this area and consequently be designed proportionately, taking into account the unintended consequences of duplicative and conflicting obligations across multiple jurisdictions. To also avoid introducing measures that would require additional data collection on minors (at odds with data minimisation principles), age verification should be restricted to 18+ only, with age assurance measures in place to facilitate age-appropriate experiences.

Whilst the most egregious content should not be available on YouTube, we recognise the role that age-verification plays in ensuring that children are not able to access content which may “*impair the[ir] physical, mental or moral development*”. We are continuously looking at ways to best create an appropriate environment for family content on YouTube, so we invest heavily in the policies, technology, and teams that help provide children and families with the best protections possible.

In respect of the measures that VSPS may be required to take under Article 28b(3)(f) of the AVMS Directive, the Code should guard against the potential for divergence as regards the age at which verification measures for content which may “*impair the physical, mental or moral development of minors*” should be imposed. Practically, we believe that the appropriate age for these purposes is 18 years old, being the accepted adult age of majority in most EU countries and in line with Article 1 of the Convention of the United Nations on the Rights of the Child. This strikes an appropriate balance between the fundamental aim of protecting minors and the principle of data minimisation under the General Data Protection Regulation (**GDPR**).

The Code should also take account of existing guidance and efforts for the protection of minors online. More particularly, the Code should reflect the Irish Data Protection Commission’s “*Fundamentals for a Child Orientated Approach to Data Protection*”, which contains guidance regarding age verification. It will also be important that the Code does not conflict with the requirement under the DSA that online platforms which are accessible to minors must put in place appropriate and proportionate measures to ensure a high level of privacy, safety, and security of minors, on their service.

To illustrate the point that industry has already come up with innovative, effective and proportionate solutions, in the absence of prescriptive regulation, we have outlined below the age assurance and age-verification methods currently employed by YouTube:

- We use a combination of age assurance and age-verification to restrict the access of users to 18+ content. Age-restricted videos are not viewable to users who are: (i) under 18 years of age, or (ii) signed out. If our systems are unable to establish that the user is above the age of 18, we will request that they provide a valid ID or credit card to verify their age. We have built our age-verification process in keeping with Google’s [Privacy and Security Principles](#).
- Our age assurance models use a combination of machine learning and data from a user’s account e.g. the watch history or the types of sites a user is searching for, as well as indicators such as the longevity of the account. We do not collect new information from users in order to run this age inference model.

Question 11: What requirements should the Code have in relation to content rating? What do you consider to be current best practice? What experiences have you had using content rating systems on platforms and do you think they have been effective? What steps could we ask VSPS to take to ensure content is rated accurately by users?

In our response to the AVMS consultation submitted on 16 April 2019, we recommended taking a principles-based and proportionate approach to the requirement for VSPS to provide user “rating

systems” given the different types of users and differences between platforms and the type of VSPS they provide.

Our experience with content rating systems on YouTube has proven effective, although we continue to learn and adapt our approach with new technologies. A principles-based, non-prescriptive approach to content rating in the Code will allow platforms like YouTube to continue to explore new approaches to content rating, although human oversight remains a key element of ensuring an age-appropriate experience across our service.

We note that in the AVMS Directive, there are several references to “users”. It is important to clarify that there are two types of “users” - “creator-users” who share and upload content, and “viewing-users” who view content. Different features are required for different types of users. The “users” to whom a VSPS should offer the ability to rate content under Article 28b(3)(g) should be a “creator-user”. On YouTube, creator-users can, and are asked to, identify, upon upload, whether a video should not be available to children under the age of 18. We combine this with our own classifiers and reviewers to establish the content that should be available only to those over the age of 18. We believe that only the “creator-user”, not the “viewing-user” should have access to such a feature, as crowd-sourced “user ratings” would be unreliable and are, in our experience, subject to abuse.

In particular, VSPS should only be required to age-restrict content at 18+. Offering age ratings (and age-gating) with greater levels of granularity is not feasible at the scale required for user-generated content uploaded to VSPS, nor would it be necessarily helpful to users watching content originating from multiple territories (from within and outside the EU).

Whilst there is a level of consistency around the level at which content is rated 18+ across different platforms and territories, more granular age ratings are also likely to be highly subjective (and culture dependant), with different viewers (and legal guardians) holding very different views on whether a piece of content is appropriate for a 13 year-old or a 15 year-old, for instance. We would also note the varying ages of digital consent adopted in different Member States, which must also be taken into account.

To ensure accurate content rating, VSPS should implement clear policies that align with conventional standards and invest in automated systems with human oversight. Regular re-evaluation of policies, flexibility in adjustments, and encouraging users to report inappropriate content are vital steps to maintaining an environment that is safe and aligned with community needs.

Question 12: What requirements should the Code have in relation to parental control features? How can we ensure that VSPS providers introduce the mechanism in a user-friendly and transparent way? Can you point to any existing example of best practice in this area? Should parental controls be ‘turned-on’ by default for accounts of minors or where age is not verified?

At YouTube, we recognise the importance of balancing child safety, children's data privacy, and their rights to access services and information and we take a multi-faceted approach to protecting children. Through YouTube Kids and our parental controls features on YouTube and Family Link, we continue to develop features that enhance the user experience. We therefore recommend maintaining flexibility in the Code regarding such safety features, maintaining a high-level approach to such requirements.

Our proactive approach towards online safety is evident in features like YouTube's supervised experience, Family Link and the default settings we have for unverified age accounts. By favouring supervised digital experiences and defaulting to safety-first settings, we ensure an added layer of

precaution. We suggest that any guidance is non-binding and used to spotlight such proactive features, using these types of mechanisms to inform best practice safety measures across digital platforms.

As an illustration of best practice, YouTube Kids showcases the advantages of age-centric platforms. YouTube Kids was built from the ground up to be a safer and simpler experience for children to explore, with tools for parents and caregivers to guide their journey. We work to identify content that is age-appropriate, adheres to our quality principles, and is diverse enough to meet the varied interests of children globally. For parents who believe their child is ready for a broader experience, we offer a [supervised experience](#). Whichever content settings the parent chooses, the child cannot gain access to 18+, age-restricted content through the account.

Safety isn't solely about control mechanisms. YouTube underscores the importance of digital literacy which is also a feature of Article 28 of the AVMS Directive. Providing parents and children with knowledge tools can support and lead to informed and responsible online conduct. It's paramount for the Code to guide platforms beyond mere control mechanisms, emphasising the significance of user education and awareness.

Question 13: What requirements should the Code contain to ensure that VSPS provide for effective media literacy measures and tools?

YouTube supports a flexible Code that emphasises effective media literacy and we believe that YouTube's best practice could be shared by CnaM through non-binding guidance to ensure that all VSPS providers champion media literacy in the digital age. This would ensure that VSPS can continue to consider new and innovative approaches to promoting media literacy rather than being constrained to focusing on specific measures promoted by the Code.

YouTube believes that promoting media literacy requires a multi-pronged approach. By emphasising authoritative sources, offering contextual information, providing transparent communication, and launching educational initiatives both on and off our platform, we seek to ensure users are not only consuming content but are also well-informed, critical thinkers. We are committed to these endeavours and continuously seek innovative ways to bolster media literacy, especially among young internet users.

A key focus of our media literacy measures has been the elevation of reliable and official information, thereby assisting users in differentiating between verified and potentially misleading content. Our fact-check feature complements our other initiatives like the Breaking News and Top News shelves which guide viewers to authoritative sources, whether they're browsing their homepage or actively searching for news topics.

To address misinformation and provide additional context, in 2018 we rolled out information panels that offer diverse contextual data. These range from linking to trusted encyclopaedic sources for debunking enduring myths (like so-called "flat earth" theories) to connecting users with authoritative health authorities such as the WHO, HSE, CDC, or local health experts in the context of evolving situations like the COVID-19 pandemic. These panels are also instrumental in challenging misinformation that emerges swiftly during fast-paced news cycles where factual uncertainties are prevalent.

Understanding the necessity for clear communication with our users regarding safety measures and available tools, YouTube launched the "[How YouTube Works](#)" website in 2020. This platform offers lucid information on our content policies, the repercussions of violating our Community Guidelines, and elucidates the tools users have at their disposal. This encompasses both privacy controls and parental controls, facilitating a custom-tailored YouTube experience for each user. Since its inception,

this platform has been a pivotal resource, arming users with the knowledge to make safer, more informed decisions on YouTube.

Beyond our platform-specific endeavours, YouTube is dedicated to fostering media literacy in broader contexts. We champion two distinct education programmes: "Be Internet Legends" and "Be Internet Citizens." The former imparts both practical and behavioural skills to schoolchildren, empowering them to traverse the internet securely. "Be Internet Citizens" aids young individuals in bolstering their critical thinking abilities online. Collectively, these initiatives have enlightened over 1.9 million children, equipping them with vital digital skills, transforming them into more discerning internet users. What amplifies the impact of these programmes is our collaboration with external experts (such as Barnardos in Ireland) and the independent assessments we conduct, ensuring they effectively alter the online behaviours of young individuals. In November 2022, we also launched our 'Hit Pause' media literacy campaign. This programme seeks to teach viewers critical media literacy skills via engaging and educational public service announcements via YouTube home feed and pre-roll ads, and on a dedicated YouTube channel. The YouTube channel hosts videos from the YouTube Trust & Safety team that explain how YouTube protects the YouTube community from misinformation and other harmful content, as well as additional campaign content that provides members of the YouTube community with the opportunity to increase critical thinking skills around identifying different manipulation tactics used to spread misinformation.

Lastly, we would note that YouTube is now also subject to transparency requirements under DSA, which requires that users are informed on how information is suggested to them or prioritised for viewing on the platform.

5.2 Terms and Conditions, Content Moderation and Complaints

Question 14: How should we ask VSPS providers to address online harms in their terms? What examples are there of best practice in relation to terms and conditions including content moderation policies and guidelines?

In our view, the focus of the Code should be on implementing the AVMS Directive and ensuring consistency with the DSA, and this should be done through a principles-based, non-prescriptive code. Best practice in relation to specific terms and conditions and transparency of terms could be provided through non-binding guidance.

YouTube's own policy framework shares common values with the safeguards of Article 28b of the AVMS Directive. Our Community Guidelines ban categories of material including hate speech, harassment and incitement to violence and do not allow content that endangers the emotional and physical wellbeing of minors. Users of our platform must follow rules of conduct, including rules against sexualisation of minors, harmful or dangerous acts involving minors, inflicting emotional distress, and cyberbullying. We provide mechanisms for users to report inappropriate content or behaviour towards children, including child endangerment.

For sanctions on users who break the rules, in most cases the first violation of our Community Guidelines will result in a warning. Then we have a general three-strikes rule where three policy violations lead to account termination, but we may also terminate the account at first offence for egregious violations. In instances when child sexual abuse material is found in user-generated content on our services, we report it to the relevant authorities and we disable the account.

Question 15: How should we ask VSPS providers to address content moderation in the Code? Are there any current practices which you consider to be best practice? How should we address automated content detection and moderation in the Code?

As practices in relation to content moderation continue to evolve, we believe the Code should not be prescriptive about specific practices and instead focus on a principles-based approach. Best practice examples could be shared through non-binding guidance and could be changed quickly as new innovative approaches to content moderation are developed. Given the scale of VSPS content and the varied approach different services take to content moderation, the Code should not limit VSPS to specific practices or require practices that would be disproportionate to implement.

YouTube uses a combination of automated and human evaluation to ensure content complies with our policies. In 2020, the most recent year for which we have figures, more than 20,000 people across the globe helped enforce Google's policies and moderate content. Our reviewer teams work around the world, 24 hours/7 days a week, speaking many different languages and are highly skilled. Our goal is to achieve both accuracy and scale in our work. Our flagging system allows our user community to notify us of any content that violates our guidelines and to help enforce our policies. Moreover, we have also developed a "Priority Flagger" program to help encourage submissions of multiple high-quality flags about content that potentially violates our Community Guidelines.

As set out in the YouTube [Transparency Report](#), between January 2023 and March 2023, we took down 8.7 million channels and 6.4 million videos on YouTube that failed to comply with our policies. Video removals resulted from approximately 6 million automated flags, 362 thousand user reports, 43 thousand organisation reports and 7 government reports. 72.3% of policy-violating YouTube videos were removed before they were viewed less than ten times. While we facilitate and encourage flags by users, in practice, low actionability rates from user flags have required significant investments in our automated systems.

We recognise that in some instances, promptness is more important than others. For instance, in cases of child sexual abuse material, YouTube uses several automated systems such as hash-matching, CSAI Match, machine learning classifiers and Content Safety API combined with human review.

Other types of potentially illegal content (such as potential terrorist and violent extremist content, hate speech, or non-consensual explicit images) either have no standard definition or require contextual understanding to determine lawfulness, such as whether the subject of the content has consented to its availability online or whether the content has an educational focus, appears as part of a documentary, or represents artistic expression. Deciding whether content is illegal is not always a determination that YouTube is able to make alone and we balance taking action against content with respect for the rights to freedom of expression and access to information.

We recognise that automated content detection technologies continue to evolve. Recent research has also shown that even small changes to images - imperceptible to the human eye - can fool computer systems into missing what is obvious to human reviewers. Measures are improving all the time, but they should only be deployed carefully, and when judged effective by individual companies based on their specific service's needs. Given this complexity and the requirement for a risk based approach, the Code must allow VSPS to innovate by refining existing technologies and exploring the development of new technologies.

Education of creators is also key¹⁰. More than 80% of creators who receive a warning never violate our policies again.

Finally we would note that content moderation procedures for larger platforms will also have to be assessed as part of the DSA risk assessments, and for which the DSA has laid out extensive

¹⁰ YouTube regularly updates Creators on changes. See for example: <https://blog.youtube/inside-youtube/an-update-to-community-guidelines-warnings/>

frameworks regarding terms and conditions disclosures, reporting mechanisms, user notice, internal appeals, out-of-court redress, and transparency. We urge CnaM to ensure there is alignment and consistency between the Code and the DSA in this area.

Question 16: What requirements should the Code include about procedures for complaint-handling and resolution, including out-of court redress or alternative-dispute resolution processes? To what extent should these requirements align with similar requirements in the DSA? What current practices could be regarded as best practice? How frequently should VSPS providers be obliged to report to the Commission on their complaint handling systems and what should those reports contain? Should there be a maximum time-period for VSPS providers to handle user complaints and if so, what should that period be?

DSA alignment

YouTube has long allowed creators to appeal against our Community Guidelines actions. Further, the DSA internal complaint mechanism mandates that online platforms provide a system where users can lodge complaints electronically and free of charge. This ensures that users have a formal channel to voice their concerns about specific decisions made by online platforms relating to information they provided. We recommend users go through our internal complaint mechanisms to seek resolution before going to out-of-court redress, which is also a possibility the DSA offers. The Code should avoid placing prescriptive obligations on VSPS and should focus on implementing the requirements of the DSA and AVMS Directive.

We believe that the DSA out-of-court dispute settlement process is broad and far-reaching; the Code should not therefore seek to set up a parallel, competing process for users to challenge the decisions of VSPS, pursuant to the AVMS Directive, in matters regulated by the DSA. In particular, users should not be offered the opportunity to re-open individual complaints that have already been escalated and decided through appropriate procedures. Any overlap between the 2 parallel regimes would lead to legal uncertainty for service providers as to their obligations under DSA and confusion for users as to which out-of-court mechanism they may have recourse to for settling disputes arising in relation to content moderation decisions. We would therefore welcome CnaM's suggestion in the Call for Inputs of an integrated complaint-handling system that covers both DSA and Code related matters, albeit the DSA regime is aimed at individual complaints, whilst the AVMS regime targets systemic issues, which inevitably suggests a higher bar.

To the extent that any matter falls outside the remit of the DSA, the AVMS Directive requires that out-of-court dispute resolution should be available in respect of VSPS failure to comply with its obligations under Article 28b(1) and (3) (i.e. systemic failures as opposed to individual cases). Any out-of-court dispute settlement mechanism should be designed in a manner that avoids fragmentation and ensures high-quality decisions.

Given the potential confusion and complexity involved with having 2 parallel out-of-court dispute settlement processes in place, and in order to avoid unmerited claims overwhelming any arbitrator or ADR provider involved, the Code should consider appropriate limitations both on the admissibility and competence of the out-of-court dispute settlement bodies. We suggest the Code establishes (1) a requirement to go through a VSPS' internal appeals process before having the right to revert to an out-of-court dispute settlement mechanism; (2) no requirement for VSPS to engage if a user triggers the out-of-court dispute settlement mechanism, but rather ability to refuse to engage in good faith (particularly where the same or similar issue has already been decided or is pending); (3) the choice of a shortlist of out-of-court dispute settlement mechanisms a user can revert to should be with VSPS to ensure centralisation and adequate expertise; and (4) any out-of-court settlement body

should be required to have a codified set of minimum standards, and CnaM should regularly audit these bodies' compliance against these standards.

Reporting on complaint handling

Regarding the frequency and the content of the reports to CnaM on complaint handling systems, we launched our first quarterly [YouTube Community Guidelines enforcement report](#) in April 2018. That report contains data on actions YouTube takes with regard to content on the platform that violates our policies. This currently includes: flagging (users and automated); video, channel, and comment removals; appeals and reinstatements; and highlighted policy verticals.

Handling user complaints

The DSA requires platforms to act in a timely, diligent and non-arbitrary manner in processing notices, taking into account the type of illegal content being notified and the urgency of taking action. The exact time frames in which this should be done were purposely not set out in the DSA. The harmonised approach of the DSA precludes Member States from laying down specific turn-around times for the removal of allegedly illegal content, recognising the need for an appropriate balancing assessment regarding the rights of affected individuals with respect to each removal or disable as specifically required under the DSA¹¹.

We would recommend that the same applies to complaint handling - service providers should be required to act in a "*timely*" manner, ensuring consistency with the DSA.

5.3 Possible Additional Measures and Other Matters

Question 17: What approach do you think the Code should take to ensuring that the safety measures we ask VSPS providers to take are accessible to people with disabilities?

We welcome the opportunity to provide information to CnaM on the measures that YouTube has in place to ensure that the service is accessible to people with disabilities. We hope that this information will assist CnaM in considering that a principles-based approach in a Code would be most appropriate to allow VSPS to be more accessible to users with disabilities.

We strive to ensure that YouTube is a platform for everyone, including users and creators with accessibility requirements. The Code should recognise people's varying needs and accessibility is an important criterion for how we develop and innovate our products; however, we do not believe the Code should take a prescriptive approach to such features and should remain principles-based enabling VSPS to innovate and deliver new accessibility features, rather than focusing on a list of features that may become outdated regularly.

At YouTube we continue to innovate to deliver new features. For example, we have recently redesigned icons on our app to be readable, consistent and clearly understood. Thanks to long term investments in machine learning, we now provide automatic captions in more languages. We also offer machine translated captions for mobile that enable viewers to translate their captions to 16 languages. YouTube has captioned over six billion videos with more than one billion users watching videos with captions enabled every day. Our app also works with Android features and informs users how to turn on or disable features that can aid with app usage.

We would also note that Article 47 of the DSA facilitates and encourages the drawing up of codes of conduct by the European Commission for the purposes of improving accessibility to people with

¹¹ Recital 22 of the DSA

disabilities. It will be important that the Code take a non-prescriptive approach so as to ensure that CnaM's approach can be adapted in line with these codes of conduct once developed.

Question 18: What approach do you think the Code should take to risk assessments and safety by design? Are there any examples you can point us towards which you consider to be best practice?

The requirements to carry out risk assessments is a feature of the DSA. Specifically, the DSA requires providers of very large online platforms (**VLOPs**) and very large online search engines (**VLOSEs**) to identify, analyse and assess systemic risks in the EU stemming from the design, functioning or use made of their services, including the risk of dissemination of illegal content through their services. The DSA only introduces such a risk assessment regime for VLOPs and VLOSEs and not for other providers of intermediary services. Given the harmonised, deliberately and carefully graduated approach of the DSA, the DSA in turn requires that Member States would not impose "VLOP/VLOSE-like" obligations on providers of intermediary services that do not qualify as such under the DSA.

In addition, Member States must not adopt any approach which would overlap with the VLOP/VLOSE risk assessment requirements of the DSA¹². To do so would result in VLOPs/VLOSEs being subject to multiple different risk assessments which seek to achieve the same overarching objective i.e. the mitigation of systemic risks stemming from the design or functioning of the service. Introducing additional risk assessment obligations on VLOP/VLOSE providers through national rules would result in a potential divergence between EU legal regimes, fragmentation of the internal market and potential legal uncertainty. These are issues that the DSA expressly wishes to avoid through its harmonisation of rules applicable to intermediary services¹³.

As such, Google urges CnaM to ensure that the Code is aligned to avoid multiple requirements for risk assessments which cut across the harmonised approach of the DSA.

Question 19: How do you think that cooperation with other regulators and bodies can help us to implement the Code for VSPS?

Regulatory input from different stakeholders with different perspectives can enrich the overall approach. However, there are also potential challenges stemming from the involvement of multiple organisations. Within Ireland, it will be important for CnaM to align its approach with other regulators, for example the Data Protection Commission.

Jurisdictional issues will need to be carefully considered in the development of a national regulatory model. For example, it will be important to ensure that there is a level playing field between service providers that are established in Ireland as against those that are under the jurisdiction of other Member States, while also respecting the AVMS Directive's country-of-origin principle.

Question 20: What approach do you think we should take in the Code to address feeds which cause harm because of the aggregate impact of the content they provide access to? Are there current practices which you consider to be best practice in this regard?

We would expect any potential risk factors arising from feeds or recommender systems to be addressed under the DSA risk assessment and risk mitigation regime. Inclusion in the Code risks cutting across the harmonised approach required by the DSA.

¹² Article 34 DSA.

¹³ Recital 9 DSA.

Question 21: Do you have any views on how requirements for commercial content arranged by a VSPS provider itself should be reflected in the Code?

As per the guidelines of the AVMS Directive¹⁴, “policies implemented by the services [should be] aimed at guaranteeing the appropriateness of the audiovisual content around or within which commercial communications of a specific third party brand would be displayed”. We believe that the Code should mirror this high-level requirement without prescriptive measures.

The concepts of “marketing, selling or arranging audiovisual commercial communications” are not clearly defined under the AVMS Directive, but cover both (a) paid advertising and (b) product placements/ sponsorships of organic (user-generated) content. A VSPS does not, as a rule, play any role in marketing, selling or arranging product placements or sponsorships of organic content. In any instances where they do so, it is appropriate that they are involved in ensuring that the relevant audiovisual commercial communication complies with the qualitative rules set out in the AVMS Directive’s Article 9(1).

In the context of paid advertisements which accompany, sit alongside or are served before or during programmes and user-generated videos (“**paid ads**”), the VSPS may play a limited, “technical” role in the marketing, selling or arranging of the paid ad on its service, depending on how those terms are interpreted.

VSPS in general have means of policing compliance with their policies or, where necessary, giving effect to regulators actions against non-compliant advertisers. However, there are several practical and operational challenges in placing responsibility on VSPS for ensuring that paid ads on its platform comply with the qualitative restrictions under Article 9(1) - which in some instances (e.g. Article 9(1)(c)(i), (iii) and (iv)) are subjective in nature and/or related to matters that are solely within the knowledge of the relevant advertiser. For instance Article 9(1) if strictly applied, would require a VSPS to make subjective judgments in relation to the nature of paid ads uploaded to its platform to determine whether it, for example, “encouraged behaviour prejudicial to health and safety”, or “grossly prejudicial to the protection of the environment”.

It is also worth noting that VSPS will be subject to advertising transparency obligations under the DSA. These include ensuring that advertisements are clearly labelled as such and providing users with information regarding advertisements presented on the service. It is likely that certain of the DSA obligations will also assist VSPS in complying with the requirements of the AVMS Directive in relation to commercial content arranged by them. As such, in line with the goals outlined in the Call for Inputs, it will be important that these DSA requirements are taken into account under the Code, so as to maximise the potential for synergies in how platforms comply with it and the DSA.

Question 22: What compliance monitoring and reporting arrangements should we include in the Code?

The AVMS Directive requires CnaM to assess the appropriateness of the measures that VSPS providers take under Article 28(b). We welcome a monitoring framework that is proportionate and focused on structural compliance rather than prescriptive reporting requirements.

We believe it is important for regulators to take a holistic, systems-focused view of compliance. When the regulator assesses the systems put in place by platforms, it should do so by primarily focusing on

¹⁴ https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=uriserv:OJ.C_.2020.223.01.0003.01.ENG&toc=OJ.C:2020:223:TOC

the relevant outcomes to be achieved, with different VSPS being given scope to determine how best those outcomes can be achieved. We see this as a more appropriate and proportionate approach to regulation, as opposed to setting out a detailed set of rigid requirements that all VSPS must meet.

Question 23: Should the Code have a transition period or transition periods for specific issues? Which areas touched on in this Call for Inputs may VSPS providers require time to transition the most? What time frame would be reasonable for a transition period?

We note that, as a general rule, transition periods are an important part of a proportionate regulatory framework, ensuring services have sufficient time to get implementation right.

While any implementation requirements will depend on the final shape of the Code, the more prescriptive the Code is, the more likely that providers will need longer transition periods to implement specific measures in response to these requirements. As outlined in our submissions, we would urge CnaM to ensure the Code remains focused on implementing the AVMS Directive and mirrors its approach of imposing requirements on VSPS without applying prescriptive and potentially overly burdensome obligations on platforms, which would further delay Ireland's effective transposition of these measures.

Given CnaM is still in the early stages of its development of the Code, it is not possible to consider the exact requirements at this stage and therefore it is difficult for any respondent to determine the specific length for such a transition period. By way of comparison, we would note that the DSA allowed a 15 month transition period for most in-scope providers to comply with its provisions. In this instance providers were also well informed of what the DSA obligations entailed well in advance of the commencement of that 15 month period.

We look forward to working with CnaM as it develops the Code to determine whether, and to what extent, such a period is required.

Headline Submission to Coimisiún na Meán on Online Safety Code for Video-Sharing Platform Services (VSPS)

September 4, 2023

This submission is led by Headline with input from people using both Shine and See Change services. It was also produced as part of an alliance with Samaritans Ireland, The National Office for Suicide Prevention, the National Suicide Research Foundation, and the Mental Health Unit of the HSE. Many of the same concerns and perspectives are raised by this group in their individual submissions. As Shine's national programme for responsible reporting and representation of mental ill health, Headline's submission is focused on our particular area of expertise: Media monitoring, stigma reduction, and suicide reporting guidelines. Also included in this submission are survey responses to questions posed by Coimisiún na Meán.

'I think I'd just like to share how vulnerable I feel young people are on social media in relation to mental health content particularly. As an adult who has been in recovery a long time, I found the content on (VSPS) particularly harmful and it took me a long while to notice the damage the content was having on my recovery. I've never managed to use (VSPS) in a way that feels safe for me and I can't imagine my younger self ever being able to make that connection.'

Survey respondent

1 Introduction

1.1 About Headline

Headline is Shine's national programme for responsible reporting and representation of mental ill health. We are a recognised global leader in innovative media and mental health practices. The media has an enormous impact on how people form their opinions about others and the society in which they live. This includes how people form their opinions on mental health and illness. Headline works collaboratively with media professionals, policymakers, academics and other stakeholders to reduce suicide contagion via the media, improve the media's ability to cover mental health stories, and enhance audiences' understanding of mental health experiences. We achieve this through our media monitoring, student and professional workshops, research, policy collaborations, [fellowships](#) and [awards](#). Most recently, Headline has provided recommendations to the World Health Organization on their suicide reporting guidelines, the Broadcasting Authority of Ireland on the Code of Programme Standards, and has presented at the World Congress on Suicide Prevention.

1.2 About Shine

Shine is a national organisation in Ireland providing information and support for people affected by mental health difficulties. Shine supports individuals and family members through individual recovery work, peer support groups, training and education and stigma reduction programmes. Shine also advocates for social change by promoting and defending the rights of all those affected by mental health difficulties to equal support and quality services. Shine is Ireland's only national mental health organisation specifically founded to support all family members. Our national programmes are See Change and Headline. [See Change](#) campaigns to end the stigma around mental health by changing public attitudes and behaviour towards people with mental health challenges for the better. [Headline](#) works towards responsible reporting and representation of mental ill health in the Irish media.

1.3 Elevating the voice of lived experience of mental ill health

In Shine, we believe that people with lived experience of mental health challenges have invaluable knowledge and can offer great insights to improve mental health support and services in Ireland, as well as other areas affecting people's mental health. In collaboration with people that use our services, Shine supports the voice of lived experience to shape and influence policy, research, legislation and mental health theory and practice in Ireland. We do this through Shine's Voice Platform and See Change's Ambassador programme.

2 Survey on Experiences of Online Video Sharing Platforms

While Coimisiún na Meán has asked for the perspectives of children and young people, we believe that harmful online video content affects all ages, especially in relation to mental health content, and content negatively affecting adults' mental health. For this reason, Shine shared survey questions with our Voice Platform participants and See Change Ambassadors. Where an individual company that provides a video sharing service as part of their platform has been named, we have substituted with 'VSPSX'. In relation to harmful content, one VSPS accounted for 50% of the responses, with the remaining 50% split between four different VSPSs. The responses are below.

2.1 Summary of survey responses

Respondents were in the age range of 18-54

Key these include:

- Negative impact on respondents' mental health due to a lack of trigger warnings
- Feelings of powerlessness when harmful content not removed/ blocked
- Fears of copycat behaviour in relation to suicide, self-harm and eating disorders
- Misinformation regarding different mental health experiences, which reinforce discrimination and stigma.

How safe do you feel on video sharing platform services (VSPS)?

37% said they feel safe online (just 11% saying they felt very safe)

What do you like about being able to watch or share videos?

'I love being able to make and share inspirational reels and posts to spread awareness of my cause (mental health stigma reduction). I love when I can watch inspirational and educational videos.'

'Good to be able to share my mental health journey and learn from others as it is real people talking.'

Are you concerned about any videos that you see on websites or on apps?

89% said they were concerned about videos they see on websites or apps.

What types of videos concern you the most?

'VSPSX videos of people very unwell with eating disorders sharing what they eat.'

'I follow a girl on VSPSX (and) VSPSX who is anorexic and dying on camera and this is extremely upsetting as I feel powerless to help.'

'There is a lot of video content that is very triggering and dangerous for vulnerable people. I believe there's a lot of content that could lead to 'copycat' behaviors.'

'Videos that contain triggering subjects with no warning, Cruelty to people or animals, hate speech or stereotypes about mental health or disability.'

Do you feel you have enough control over the type of videos you see on websites or apps?

63% said no.

'From time to time I will see these videos and click not interested so they stop but after a few weeks they come back around.'

Do you think that companies who run websites or apps that allow videos to be watched or shared should do anything to make things safer for you or your friends or family?

89% said yes.

'When a video is reported its hard to get it removed even though it genuinely is harmful. There needs to be a better procedure.'

'I agree with freedom of expression so it's a tough one but I think platforms should be doing something to protect the users and also to protect the people making the videos (especially when they're in very vulnerable situations and are too unwell to know that what they're sharing may harm themselves and others).'

'They could listen and respond to feedback (this rarely happens), moderate videos and comments, issue warning to repeat offenders.'

'Reviewing the content uploaded to their platforms helps as they can apply warnings before the videos which is often accompanied by a blur so that you can make the choice on whether or not to view the content.'

53% of respondents had previously reported concerns about video content to the company in charge of the website or app.

Survey respondents were asked for their recommendations on making online video sharing safer. Here are some of their responses:

- *'If videos are reported the user who posted should be notified that it was found harmful, even if the bot thinks it isn't.'*
- *'People have to have identity verified before setting up account so can be held responsible for their actions. Give users complete control over being able to avoid video subjects that may trigger upset for them.'*
- *'I think we need more education in the public for example similar to the media guidelines that headline has a similar education campaign around social'*

media to highlight to people to think about vulnerable populations who may be viewing their content. Also I think that the platforms could have clear suggestions on kinds of content that users should think twice about posting.'

- *'Actionable Consequences. Penalties to be imposed on companies that allow inappropriate video sharing.'*
- *'Increase content review jobs to keep up with the ever-growing wealth of content. Shadow ban harmful content without infringing on the right to free speech.'*

3 Response to Coimisiún na Meán Questions

Question 1: What do you think our main priorities and objectives should be in the first binding Online Safety Code for VSPS? What are the main online harms² you would like to see it address and why?

As a mental health organisation, Shine prioritises online harms relating to mental health, mental illness, and suicide. While the Online Safety Codes will be required to cover an enormous range of potential and established harms, we have identified priorities of most concern in our field of work.

- I. Ensure VSPS are minimising harmful content to all age ranges, while maximising effective opportunities for help and support. Harmful content affects all ages, especially content that poses a threat to life. The sharing of suicide, self-harm, and eating disorder methods should be treated as an immediate priority for the Code.
- II. The Code must ensure VSPS support content moderators and have a minimum standard of care for their content reviewers. If CnaM also intend to develop a 'spot-checking' apparatus, the same minimum standard of care should be applied to CnaM staff involved in harmful content review.
- III. Hold platforms accountable through effective evaluation and monitoring of complaints and reports, made publicly available. Build in positive reinforcement mechanisms for VSPS' with good compliance.

Since 2007, Headline has been monitoring Irish news media's performance in relation to established suicide and mental health reporting guidelines, namely those from Samaritans, the World Health Organization, Bodywhys, UK-based mental health charity, Mind, and Mindframe in Australia. While based on an enormous body or

research, these guidelines are non-binding. Headline has no power to enforce these guidelines and we rely heavily on stakeholder engagement, education programmes, and additional supports for media working in this area. Harmful content has been addressed in professional media for a long time, and many apparatuses are in place to support audiences make informed decisions about the content they consume. We welcome and support any effort by Coimisiún na Meán to build robust procedures and codes that give the same controls to audiences of social media.

Question 2: What types of online harms do you think should attract the most stringent risk mitigation measures by VSPS? How could we evaluate the impact of different types of harms e.g. severity, speed at which harm may be caused? Is there a way of classifying harmful content that you consider it would be useful for us to use?

All online harms relating to self-harm or suicide should be managed with the highest priority regardless of the intent of the post. Recent social media 'games', and harmful content emerging in the aftermath of a high-profile suicide should be treated with urgency, with risk mitigation measures that support public health concerns. Headline has observed challenges to this in professional media when there is a conflict of rights, namely in the sharing of details from a court-case. Headline assumes that all social media users and 'uploaders' will be treated the same, regardless of their professional status. Therefore, professional media using social media to share content related to these harms must also be limited, while also following professional codes, e.g. Press Council's Code of Practice, and Coimisiún na Meán's Code of Programme Standards. Media platforms, whether they be traditional formats or online, should endeavor to follow the same standard of content across all platforms.

'I love the creativity and freedom the video sharing platforms have allowed for, but there are a lot of unsafe areas out there, that we need to tackle. The internet needs to be as safe as the streets are, hence more rules and regulations need to be put into place.'

Survey respondent

Headline has over a decade's experience influencing Ireland's media sector and many lessons have been learned in that time. Together with our partners across Ireland's Connecting for Life strategy, Headline would be interested in speaking further with Coimisiún na Meán on establishing thresholds for harm, which should be co-

designed with social media users of all ages as well as mental health and industry experts.

Question 4: What approach do you think we should take to the level of detail in the Code? What role could non-binding guidance play in supplementing the Code?

From our experience, it is important to include as much detail as possible when writing codes. For example, an existing non-binding code in broadcasting instructs programme makers to include helplines on all programmes that may cause distress to audiences. In recent years, Headline has identified a trend where broadcasters provide a link to a website containing a multitude of helplines, rather than providing immediate assistance to the distressed person by showing a number on the screen or calling a number out loud. In calling out a web address, rather than a helpline, the broadcaster is making an assumption about the distressed audience's capacity, access, and means. While the codes must be robust, they must also provide enough detail to the audience, the uploader, and the VSPS to avoid misinterpretation of the intended protections. This is especially critical for harmful content that poses a threat to life.

Question 13: What requirements should the Code contain to ensure that VSPS provide for effective media literacy measures and tools?

Audience education on potential harms is crucial in building an informed and media literate population. While some VSPSs have mechanisms in place to challenge misinformation, through user-led moderation, this is not uniform across the sector and Headline would welcome a wider adoption of these mechanisms. People using Shine services, including those in See Change, have indicated the powerful impact the sharing of experiences and mental health recovery journeys has had on their mental health. Headline runs workshops with people who wish to tell their stories and create online content around mental health experiences. We created these workshops in response to the unintentional sharing of harmful details and misinformation by mental health advocates online. Similar measures could be rolled out by VSPSs or CnaM to improve media literacy among audiences and content creators alike.

Question 15: How should we ask VSPS providers to address content moderation in the Code? Are there any current practices which you consider to be best practice? How should we address automated content detection and moderation in the Code?

There is clear frustration among survey respondents about the quality and responsiveness of content moderation. While automated content detection is a useful tool to support VSPSs, the Code should reflect the urgency which VSPSs must acknowledge AND address flagged content. If a social media user has identified harmful content that poses a threat to life, automated detection should move to immediately block that content until it can be reviewed by a VSPS moderator. If a moderator chooses to allow that content, the social media user must have some recourse to alert CnaM. If CnaM finds there is a track record of moderator assessment error, there must be actionable consequences for that VSPS.

It is important to note, however, that moderator assessment error may not be the result of poor judgement, but rather as a consequence of moderator work itself. In Headline, we have robust procedures, limits and supports in place to mitigate the risks associated with over exposure to harmful content. Some of those risks include compassion fatigue, vicarious trauma, disturbed sleep, intrusive thoughts, and other mental health consequences to harmful content exposure. These measures were introduced following consultation with Headline staff who expressed concern regarding these issues. The Code should instruct all VSPSs to have similar robust procedures, including the development of staff and contractor no-risk feedback mechanisms, to support the psychosocial risks of this work. Coimisiún na Meán should also introduce a mechanism whereby contractors engaged in this work can report VSPSs for failure to comply with safe moderation practices. This may be done in collaboration with the Health and Safety Authority. Headline welcomes any opportunity to discuss these protections further.

Question 17: What approach do you think the Code should take to ensuring that the safety measures we ask VSPS providers to take are accessible to people with disabilities?

With respect to psychosocial disabilities, it is important that the Code empowers those who have identified triggers individual to their mental health condition, block content that could harm them. Many of the survey respondents who identified as having mental health conditions spoke about their concern for themes or 'triggers' that, over the course of a few weeks, or after software updates, reappeared in their social media feeds. The Code must ensure VSPSs support users autonomy when choosing content they have identified as being detrimental to their mental health. Accessibility to reporting practices must also be ensured through clarity of language and simplified reporting practices.



**Irish Heart
Foundation**

**Submission to Coimisiún na
Meán**

**Call For Inputs: Online Safety
Developing Ireland's First
Binding Online Safety Code for
Video-Sharing Platform
Services**

September 2023



Background and Introduction

The Irish Heart Foundation (IHF) welcomes the opportunity to make a submission to inform a future consultation by Coimisiún na Meán (the “Commission”) on a draft Online Safety Code Video-Sharing Platform Services.

The Irish Heart Foundation (IHF) promotes policy changes that reduce premature death and disability from cardiovascular disease (CVD). A number of the risk factors for CVD have been shown to be influenced by developments in the digital world. The rapid evolution of online platform capabilities and the sophistication of new forms of commercial communication has sparked the need for concrete action to be taken to protect children from exploitation and harms.

The Irish Heart Foundation sees an important role for the regulation of harmful content in protecting children’s health and protecting them from privacy risks, loss of reputation, commercial exploitation of personal data, profiling and cyber harassment. Today’s youth – in the womb through to adolescence - are at the epicentre of an exploding digital media and marketing landscape. Indeed, there is significant scope for the Media Commission to recognise and support the position that children hold in the digital ecosystem, as articulated by UNICEF: “that of rights holders, entitled to be protected from violations of their privacy and deserving an Internet free from manipulative and exploitative practices.”

Due to the current complexity of the regulatory framework on commercial communications – which covers media law, consumer protection law, e-commerce law and data protection law – policy makers and legislators are being faced with increasing difficulties in how to provide accountability mechanisms, and regulate for, commercial communications that appear across various platforms (traditional media and internet content).

Few would argue against the fact that there are significant disparities in whether and how online content is regulated. Issues related to misleading political advertising, ‘fake news’ and bullying have, to date, been particular areas of focus in respect of social media platforms and resultant calls for regulation. However, commercial communications also strongly influence what young people eat and drink, harming their health, well-being, and rights. Additionally, these commercial communications are incompatible with a vision for health-promoting and sustainable food systems and, as such, must be addressed by the Commission in the development of its Online Safety Codes.

How the Submission is structured

The Call for Inputs document set out a number of issues and questions, exploring a wide range of topics, many of which are outside the direct expertise of the IHF. Therefore, questions relevant to the work of the IHF, as well as groups such as the Children’s Rights Alliance of which the IHF is a member, are addressed in order. Some responses cover multiple questions, given some of the related content and to avoid duplication of responses.

Summary of Recommendations:

1. Harmful Commercial Communications, particularly marketing of high fat, sugar and salt foods and breastmilk substitutes, that infringe on fundamental rights as enshrined in the Convention on the Rights of the Child should be addressed.
2. The heightened risks of, and harms associated with, commercial exploitation and negative impact on development and health that can occur as a result of marketing practices of HFSS food and drink and Breastmilk Substitutes, must be addressed in the Online Safety Codes.
3. The Online Safety Code must be prescriptive and high-level.
4. The Online Safety Code must ensure that children are protected effectively from harmful marketing and that their Convention on the Rights of the Child rights are upheld. This includes addressing commercial communications for mixed audiences, in order to capture all the marketing that children are exposed to.
5. Online Safety Codes should protect all children, not just those old enough to have digital access.
6. The Commission should be able to assess the effectiveness of procedural measures against a set of statutory objectives that go beyond simplistic content-related benchmarks such as removal rates and response times.
7. The Commission should have the power to demand any type of granular information that is necessary for it to fulfil its supervisory tasks. Shifting scrutiny towards these processes would help address some of the causal factors that give rise to harmful content online.
8. Strong, proactive enforcement mechanisms are needed, which would apply stronger punitive measures for instances of noncompliance.
9. Child rights impact assessment (CRIA) should be mandated
10. A dedicated function within the Media Commission should relate to online harms as they relate to data protection. As recommended by the Data Protection Commission, online harms that relate to data protection should be dealt with by the Media Commission.
11. Self-regulatory bodies should not be involved in the regulation of commercial communications or in the implementation of the Online Safety Code for VSPs
12. Enforcement mechanisms should be both reactive and proactive, meaning that they should be open to both receiving notification of infringements, and detecting infringements through screenings and ongoing monitoring.

13. Continuous monitoring and enforcement mechanisms should be established (including a complaints procedure available to those with a legitimate complaint)
14. There should be a clear authority to enforce the restrictions
15. Regulated entities should not just be required to “provide periodic reports on their compliance or otherwise with codes”, but should also be forced to provide any type of granular information to the Commission that is necessary for it to fulfil its supervisory tasks
16. Provision should be made to enable independent public interest research, based on data from platforms

Question 1: What do you think our main priorities and objectives should be in the first binding Online Safety Code for VSPS? What are the main online harms you would like to see it address and why?

Recommendations:

Harmful Commercial Communications, particularly marketing of high fat, sugar and salt foods and breastmilk substitutes, that infringe on fundamental rights as enshrined in the Convention on the Rights of the Child should be addressed.

The heightened risks of, and harms associated with, commercial exploitation and negative impact on development and health that can occur as a result of marketing practices of HFSS food and drink and Breastmilk Substitutes, must be addressed in the Online Safety Codes.

Priorities and Objectives:

Given the harmful impact of food marketing is a function of both exposure and power, the objectives regulating commercial communications should include protecting health and children's rights by reducing both the exposure of children to, and power of, marketing of HFSS foods. Taking such an approach will ensure that the best interests of the child are upheld as a primary consideration and will offer children protection from the harms created by commercial communications.

The main objectives should be:

1. To protect children from the harm associated with the marketing of nutritionally poor food.
2. To provide a binding basis for a high level of public health and protection in relation to commercial communications
3. To protect the fundamental rights and freedoms of children and in particular their right to the enjoyment of the highest attainable standard of health, right to food and right to privacy
4. To uphold the best interests of the child as a primary consideration.

Commercial Communications impact on children, their health and their rights

It has been recommended that the notions of 'online safety' and 'online harms' should be defined broadly to include concerns related to digital marketing and data protection and privacy. Moreover, harmful digital marketing should be identified as a safety risk for children by States and by business actors themselves.¹

The 2020 WHO- UNICEF-Lancet Commission on the future for the world's children noted that "commercial marketing of products that are harmful to children represents one of the most underappreciated risks to their health and wellbeing".²

¹ Garde, A. et al. (2019). General Comment submission Children's rights in relation to the digital environment. [Online] Available from: https://www.ohchr.org/EN/HRBodies/CRC/Pages/Submissions_Concept_GC_Digital_Environment.aspx

² Clark, H., Coll-Seck, A.M., Banerjee, A., Peterson, S., Dalglish, S.L., Ameratunga, S. et al. (2020). A future for the world's children? A WHO–UNICEF–Lancet Commission. *Lancet* 2020; 395: 605–58. [Online] Available from: [https://www.thelancet.com/journals/lancet/article/PIIS0140-6736\(19\)32540-1/fulltext#articleInformation](https://www.thelancet.com/journals/lancet/article/PIIS0140-6736(19)32540-1/fulltext#articleInformation)

The commercial advertising and marketing of several products, services and brands are associated with poor health. Harmful commodities include but are not limited to unhealthy food and beverages, alcohol, drugs, tobacco, e-cigarettes and breastmilk substitutes. This Online Safety Code should specifically regulate harmful commercial advertising and marketing to prevent children's exposure to such advertising and marketing. Such regulation relating to the digital environment should in no circumstance be less effective than regulation in the offline environment.

HFSS Foods

Marketing of unhealthy food in digital media has been argued to harm several of the rights enshrined in the Convention on the Rights of the Child, including the rights to health, adequate and nutritious food, privacy, and freedom from exploitation.³ Evidence is emerging to suggest that concern about the public health implications of young people's exposure to digital marketing for unhealthy foods and beverages is justified.⁴

Foods high in fat, sugar and salt (HFSS) are a lead contributing factor to the burgeoning obesity crisis. This obesity crisis has major public health implications and is responsible for a considerable burden of health, social and economic harm at individual, family and societal levels. The proliferation of digital food and beverage marketing has led to concerns about the influence of this type of exposure on the health and wellbeing of children⁵, particularly given their cognitive and developmental vulnerabilities.

Digital media advertising has changed dramatically over time and is predicted to account for 60% of global advertising expenditure by 2025.⁶ A 2023 report from UNICEF and the WHO highlights that as marketing communication techniques have moved away from one-size-fits-all spot advertisements towards strategies for fostering engagement, children are now not just passive viewers of commercial messages, but rather active practitioners in the commercial communications and marketing.⁷ With the proliferation of advertising content, focussed on HFSS food and drinks, being targeted and accessible by children on online platforms without any regulation, there is an undeniable danger that the nutritional

³ Tatlow-Golden, M. and Garde, A. (2020). Digital food marketing to children: Exploitation, surveillance and rights violations. *Global Food Security* 27 (2020) 100423 <https://www.sciencedirect.com/science/article/pii/S2211912420300778> [Open Access]

⁴ Buchanan L, Kelly B, Yeatman H, Kariippanon K. The Effects of Digital Marketing of Unhealthy Commodities on Young People: A Systematic Review. *Nutrients*. 2018 Jan;10(2): 148

⁵ WHO Regional Office for Europe. (2016) Tackling food marketing to children in a digital world: trans-disciplinary perspectives. [Online] Available from: <https://www.euro.who.int/en/health-topics/disease-prevention/nutrition/publications/2016/tackling-food-marketing-to-children-in-a-digital-world-trans-disciplinary-perspectives.-childrens-rights,-evidence-of-impact,-methodological-challenges,-regulatory-options-and-policy-implications-for-the-who-european-region-2016>

⁶ WHO. (2022). Understanding the digital media ecosystem. How the evolution of the digital marketing ecosystem impacts tobacco, alcohol and unhealthy food marketing. Copenhagen: WHO Regional Office for Europe; 2022 [Online] Available from: <https://apps.who.int/iris/handle/10665/355277>

⁷ UNICEF and WHO. (2023). Taking action to protect children from the harmful impact of food marketing: a child rights-based approach. Geneva: World Health Organization and the United Nations Children's Fund (UNICEF). [Online] Available from: <https://www.unicef.org/media/142621/file/UNICEF-WHO%20Toolkit%20to%20Protect%20Children%20from%20the%20Harmful%20Impact%20of%20Food%20Marketing.pdf>

consumption habits of children and adolescents are being affected in such a way that would have the effect of endangering long-term health.

Commercial Milk Formula (CMF)

Awareness is growing of the harm of products marketed to adults for use by children. For example, inappropriate use of commercial milk formula is associated with obesity, and increased risk of diabetes and other noncommunicable diseases.⁸

The marketing of CMFs “comprehensively undermines access to objective information and support related to feeding of infants and young children. Additionally, CMF marketing seeks to influence normative beliefs, values, and political and business approaches to establish environments that favour CMF uptake and sales. In so doing, CMF marketing contributes to reduced global breastfeeding practices.”⁹ Digital platforms substantially extend the influence of marketing while circumventing the International Code of Marketing of Breastmilk Substitutes.

E-Cigarettes

The EU’s scientific committee on health, environmental and emerging risks (SCHEER), concluded that there is moderate evidence for risks of long-term systemic effects from e-cigarette use on the cardiovascular system while there is strong evidence for risks of poisoning and injuries due to burns and explosion¹⁰.

Evidence suggests that exposure to e-cigarette adverts reduces children’s perceptions of the harm of e-cigarettes and occasional tobacco smoking¹¹. Moreover, analysis shows that most vaping content on social media sites, such as Instagram¹² or TikTok¹³, which is largely used by young people, is predominantly pro-vaping. According to a 2019 study of youth exposure to e-cigarette advertising, including online and on social media channels, exposure to e-cigarette advertising was associated with an increase in subsequent past 30-days use of e-cigarette among youths and young adults. The researchers concluded that restricting

⁸ Clark, H., Coll-Seck, A.M., Banerjee, A., Peterson, S., Dalglish, S.L., Ameratunga, S. *et al.* (2020). A future for the world’s children? A WHO–UNICEF–Lancet Commission. *Lancet* 2020; 395: 605–58. [Online] Available from: [https://www.thelancet.com/journals/lancet/article/PIIS0140-6736\(19\)32540-1/fulltext#articleInformation](https://www.thelancet.com/journals/lancet/article/PIIS0140-6736(19)32540-1/fulltext#articleInformation)

⁹ Rollins, Nigel *et al.* (2023). Marketing of commercial milk formula: a system to capture parents, communities, science, and policy. *The Lancet*, Volume 401, Issue 10375, 486 – 502 [https://www.thelancet.com/journals/lancet/article/PIIS0140-6736\(22\)01931-6/fulltext](https://www.thelancet.com/journals/lancet/article/PIIS0140-6736(22)01931-6/fulltext)

¹⁰ European Commission. (2021). Scientific committee on health, environmental, and emerging risks SCHEER. Opinion on electronic cigarettes. Available here: https://health.ec.europa.eu/system/files/2022-08/scheer_o_017.pdf

¹¹ Vasiljevic M, St John Wallis A, Codling S, *et al.* (2018). E-cigarette adverts and children’s perceptions of tobacco smoking harms: an experimental study and meta-analysis. *BMJ Open* 2018;**8**:e020247. doi: 10.1136/bmjopen-2017-020247

¹² Gao Y, Xie Z, Sun L, Xu C, Li D. (2020) Electronic Cigarette-Related Contents on Instagram: Observational Study and Exploratory Analysis. *JMIR Public Health Surveill.* 2020 Nov 5;6(4):e21963. doi: 10.2196/21963. PMID: 33151157; PMCID: PMC7677028.

¹³ Perez, Sarah. (2021). TikTok is being used by vape sellers marketing to teens. Tech Crunch. Available here: https://techcrunch.com/2021/01/26/tiktok-is-being-used-by-vape-sellers-marketing-to-teens/?guccounter=1&guce_referrer=aHR0cHM6Ly93d3cuZ29vZ2x1LnNvbS8&guce_referrer_sig=AQAAAGf8maHYpCSba8a3-IS13JYVrATHjn8X0asDF0Db93DQ4GAQ9ceW4PKgGbSGPrDHO5WSCzWOo_4vOKX4QvJJDz_IPoSWWSr0dn7j2miMeu3hJVGd1zx8fhOCImNNahH7VUzTc9hy42kYX312qMLiVmEPp7ddMSO7xN9T7S77Q62T

advertising of e-cigarettes targeted at youths and young adults may reduce this cohort's likelihood of e-cigarette use¹⁴.

Similar findings were found in a 2022 study that found that this advertising (and peer influence) was significantly associated with e-cigarette initiation¹⁵. A 2021 study of social media and e-cigarette use among US youths found that youth respondents with more social media use were more likely to be exposed to e-cigarette advertisements, which led to lower e-cigarette risk perception and thus increased subsequent use. The researchers concluded that social media use among young people is associated with increased e-cigarette use through online e-cigarette advertisement exposure and subsequent decreased e-cigarette risk perception¹⁶. Meanwhile, in Ireland, a 2022 study by transition year students as part of the BT Young Scientist Exhibition concluded that social media does influence teens to vape. According to their findings from a survey of 2,000 students from across 728 secondary schools, 35% of respondents were lifetime e-cigarette users and 55% of these lifetime vapers cited being tempted or curious to try e-cigarettes after hearing a content creator or influencer talk about e-cigarettes. Similarly, 30% of all respondents were tempted or curious to try e-cigarettes when they witnessed a content creator or influencer using an e-cigarette on social media¹⁷.

Children's Rights and Online Harms

The Call for Inputs notes at page 5:

“... we want the Code to protect children and the general public from online harms while upholding and promoting human rights, including the right to Freedom of Expression.”

The Irish Heart Foundation echoes the calls from the WHO and UNICEF that the best way to respect, protect and fulfil children's rights when it comes to protecting them from harmful commercial communications is to adopt a mandatory, comprehensive approach, while recognising that steps taken to restrict these harms must integrate both a public health lens and a child rights lens.¹⁸

¹⁴ Do VV, Nyman AL, Kim Y, Emery SL, Weaver SR, Huang J. Association between E-Cigarette Advertising Exposure and Use of E-Cigarettes among a Cohort of U.S. Youth and Young Adults. *Int J Environ Res Public Health*. 2022 Oct 3;19(19):12640. doi: 10.3390/ijerph191912640. PMID: 36231939; PMCID: PMC9566774

¹⁵ Wang Y, Duan Z, Weaver SR, et al. Association of e-Cigarette Advertising, Parental Influence, and Peer Influence With US Adolescent e-Cigarette Use. *JAMA Netw Open*. 2022;5(9):e2233938. doi:10.1001/jamanetworkopen.2022.33938

¹⁶ Xia Zheng, Wenbo Li, Su-Wei Wong, Hsien-Chang Lin, Social media and E-cigarette use among US youth: Longitudinal evidence on the role of online advertisement exposure and risk perception, *Addictive Behaviors*, Volume 119, 2021, 106916, ISSN 0306-4603, <https://doi.org/10.1016/j.addbeh.2021.106916>. (<https://www.sciencedirect.com/science/article/pii/S0306460321001015>)

¹⁷ Health Research Board. (2023). Does Social Media influence teens to vape. Available here:

<https://www.hrb.ie/news/news-story/article/does-social-media-influence-teens-to-vape/#:~:text=55%25%20of%20lifetime%20vapers%20cited,e%2Dcigarette%20on%20social%20media>.

¹⁸ UNICEF and WHO. (2023). Taking action to protect children from the harmful impact of food marketing: a child rights-based approach. Geneva: World Health Organization and the United Nations Children's Fund (UNICEF). [Online] Available from: <https://www.unicef.org/media/142621/file/UNICEF-WHO%20Toolkit%20to%20Protect%20Children%20from%20the%20Harmful%20Impact%20of%20Food%20Marketing.pdf>

Online Safety Codes should also build on the global initiatives currently underway at WHO, UNICEF, and other international agencies, and should be grounded in the fundamental rights of children. Enabling children *of all ages* to achieve their full developmental potential is a human right and a critical foundation for sustainable development. Children's rights, including their rights to health, adequate and nutritious food, privacy, and to be free from exploitation, are threatened by commercial communications and their associated harms.

Any regulatory scheme should be explicitly rooted in the international human rights framework. Children's rights are enshrined in the UN Convention on the Rights of the Child (UNCRC), which was adopted unanimously by the United Nations General Assembly in 1989 and signed up to by Ireland in 1992. It has since become the most rapidly and widely ratified human rights treaty in history, and its operationalisation is supported by a series of Optional Protocols and General Comments.

Children's digital rights have been an explicit concern of the international children's rights community. Accordingly, potential infringements to such rights must sit at the heart of considerations on online harms. In their submission to the UN Committee on the Rights of the Child General Comment on children's rights in relation to the digital environment, leading academics and experts in the area of law, child development, childhood studies, psychology, food and nutrition, media studies, and child, consumer and digital rights called for the recognition of the far-reaching harms caused by digital marketing and the personal data extraction on which it is predicated, and the need to protect children from these. This, they note, is because digital media marketing is subjecting children to intense commercial practices of implicit influence, neuromarketing, attitudinal structuring and behavioural modification, without independent evaluation to ensure they do no harm. As a result "children are thus commercial digital test subjects for marketing practices affecting their development, health and privacy."¹⁹

¹⁹ Tatlow-Golden, M. and Garde, A. (2020). Digital food marketing to children: Exploitation, surveillance and rights violations. *Global Food Security* 27 (2020) 100423 <https://www.sciencedirect.com/science/article/pii/S2211912420300778> [Open Access] p1

Question 3: Do you have reports, academic studies or other relevant independent research that would support your views?

- Calvert, E. (2021) **Food Marketing to Children Needs Rules with Teeth, A snapshot report about how self-regulation fails to prevent unhealthy foods to be marketed to children.** Bruxelles: BEUC, pp. 1–24. [Download here](#)
- Chester, J., Montgomery, K. C. and Kopp, K. (2021) **Big Food, Big Tech, and the Global Childhood Obesity Pandemic.** Washington: The Center for Digital Democracy (CDD), pp. 1–72. [Download here](#)
- Clark, H. et al. (2020) **‘A future for the world’s children? A WHO–UNICEF–Lancet Commission’,** The Lancet, 395(10224), pp. 605–658. [Download here](#)
- Escalon, H. et al. (2021) **Exposure of French Children and Adolescents to Advertising for Foods High in Fat, Sugar or Salt.** Nutrients. [Download here](#)
- European Commission. Directorate General for Health and Food Safety (2021) **Study on the exposure of children to linear, non-linear and online marketing of foods high in fat, salt or sugar: final report.** LU: Publications Office. [Download here](#)
- foodwatch (2021) **Marktstudie: Fast alle Kinderlebensmittel sind ungesund.** [Download here](#)
- Garde, A. et al. (2018) **A Child Rights-Based Approach to Food Marketing: A Guide for Policy Makers.** Geneva: United Nations Children’s Fund (UNICEF), pp. 1–84. [Download here](#)
- Signal, L.N., Stanley, J., Smith, M. et al.(2017) **Children’s everyday exposure to food marketing: an objective analysis using wearable cameras.** Int J Behav Nutr Phys Act 14, 137. [Download here](#)
- Tatlow-Golden, M., Tracey, L. and Dolphin, L. (2016) **Who’s feeding the kids online? Digital Food Marketing and Children in Ireland.** Dublin: Irish Heart Foundation, pp. 1–68. [Download here](#)
- United Nations Children’s Fund and United Nations Special Rapporteur on the Right to Food (2019) **Protecting Children’s Right to a Healthy Food Environment.** Geneva: UNICEF and United Nations Human Rights Council, pp. 1–28. [Download here](#)
- WHO European Office for the Prevention and Control of and Noncommunicable Diseases (2021) **Digital food environments - Factsheet.** Copenhagen, pp. 1–14. [Download here](#)
- WHO European Office for the Prevention and Control of Noncommunicable Diseases (NCD Office) (2018) **Monitoring and restricting digital marketing of unhealthy products to children and adolescents.** Moscow: World Health Organization, pp. 1–85. [Download here](#)
- World Health Organization (2010) **Set of recommendations on the marketing of foods and non-alcoholic beverages to children.** Geneva: World Health Organisation, pp. 1–14. [Download here](#)
- World Health Organization (2012) **A framework for implementing the set of recommendations on the marketing of foods and non-alcoholic beverages to children.** Geneva: World Health Organization, pp. 1–61. [Download here](#)

- World Health Organization. Regional Office for Europe (2015) **Nutrient Profile Model**. [Download here](#)
- World Health Organization (2018) **Evaluating Implementation Of The Who Set Of Recommendations On The Marketing Of Foods And Non Alcoholic Beverages To Children - Progress, challenges and guidance for next steps in the WHO European Region**. Copenhagen: World Health Organization, pp. 1–35. [Download here](#)
- World Health Organization. Regional Office for Europe (2016) **Tackling food marketing to children in a digital world: trans-disciplinary perspectives: children's rights, evidence of impact, methodological challenges, regulatory options and policy implications for the WHO European Region**. Copenhagen: World Health Organization. Regional Office for Europe, pp. 1–52. [Download here](#)
- World Obesity Federation (2021) **Digital Deception - The Marketing of Unhealthy Food: Building a Youth-Led Response**. London: World Obesity Federation, pp. 1–10. [Download here](#)

Question 4: What approach do you think we should take to the level of detail in the Code? What role could non-binding guidance play in supplementing the Code?

Question 6: How should we design the Code to minimise the potential for conflict and maximise the potential for synergies in how platforms comply with it and the DSA?

Recommendation:

The Online Safety Code must be prescriptive and high-level.

EU legislation such as the General Data Protection Regulation, the Digital Services Act, as well as the Audiovisual Media Services Directive (the transposition of which is the basis for the development of this Online Safety Code) contain specific provisions related to child protection but most of them are principle-based and not concrete enough to be effective in practice without lengthy and costly litigation. Evidence shows that some major companies which are present in many children's lives are not sufficiently protecting them from online harms.²⁰

The new Digital Services Act contains provisions on protection of minors which are a step in the right direction. However, these only apply to online platforms and in some cases to very large online platforms (VLOPs). In addition, they fall short of prohibiting tracking and profiling minors. Article 28 requires 'appropriate and proportionate measures to ensure a high level of privacy, safety, and security' and a prohibition of displaying ads based on profiling using data from minors. However, it remains to be seen how platforms will effectively do this in practice in view of the absence of concrete legal provisions on how to operationalise these requirements.

The Audiovisual Media Services Directive (AVMSD) contains vague rules to protect minors from inappropriate on-demand media audiovisual services. These include 'encouraging' Member States to ensure that self-and co-regulatory codes of conduct are used to effectively limit the exposure of children and minors to audiovisual commercial communications for alcoholic beverages (Recital 11) or it being necessary to set out 'proportionate rules' on protecting minors from harmful content (Recital 26), or to take 'appropriate measures to protect minors from content that may impair their physical, mental or moral development' (Recital 28). Article 12 states that programmes 'which may impair the physical, mental or moral development of minors are only made available in such a way as to ensure that minors will not normally hear or see them' yet without giving any specifics. Similar provisions apply under the Article 28a to video-sharing platforms.

The UN Committee on the Rights of the Child in its General Comment no.25 on children's rights in relation to the digital environment state that States should require the business sector to undertake children's rights due diligence and child rights impact assessments and disclose them to the public with consideration of the 'severe impacts of the digital

²⁰ See, for example, BEUC action on TikTok: BEUC (2023). Holding TikTok accountable – a reality check: Letter to ERGA.[Online] Available from: <https://www.beuc.eu/letters/holding-tiktok-accountable-reality-check-letter-erga>

environment on children.²¹ The UN Committee also state that States should require all businesses that affect children's rights in relation to the digital environment to implement regulatory codes and frameworks to adhere to the highest levels of privacy and safety standards and encourage them to take accountability and measures to innovate in the best interests of the child.²²

Looking at existing legislative provisions and lack of detail in their implementation, it is clear that no decisive approach currently exists to protect minors from harms to children caused by commercial communications. Therefore, the Code must be prescriptive and high-level.

Reliance on the development of codes of conduct that are not legally enforceable or subject to sanctions for non-compliance will not be sufficient. We know that children's digital media choices and data control possibilities are shaped by the design and functionalities of communication spaces, control of which rests neither with them, their parents or indeed national regulators.²³

²¹ UN Committee on the Rights of the Child, General Comment no 25 (2021) on children's rights in relation to the digital environment, CRC/C/GC/25, para 38 Taken from Children's Rights Alliance Submission to Coimisiún na Meán Call For Inputs on Developing First Online Safety Code 2023

²² UN Committee on the Rights of the Child, General Comment no 25 (2021) on children's rights in relation to the digital environment, CRC/C/GC/25, para 39. Taken from Children's Rights Alliance Submission to Coimisiún na Meán Call For Inputs on Developing First Online Safety Code 2023

²³ Macenaite, M. (2017). From universal towards child-specific protection of the right to privacy online: Dilemmas in the EU General Data Protection Regulation. *New Media & Society*, 19(5), 765–779. [Online] Available from: <https://doi.org/10.1177/1461444816686327>

Question 8: How should we ask VSPS providers to introduce a feature that allows users to declare when videos contain advertising or other type of commercial communications? Should the Code include specific requirements about the form in which the declaration should take? What current examples are there that you regard as best practice?

The Council of Europe has recommended that ‘States should take measures to ensure that children are protected from commercial exploitation in the digital environment, including exposure to age inappropriate forms of advertising and marketing.’²⁴

The UN Committee on the Rights of the Child has reiterated this in their recent General Comment and has recommended that:

“States parties should make the best interests of the child a primary consideration when regulating advertising and marketing addressed to and accessible to children. Sponsorship, product placement and all other forms of commercially driven content should be clearly distinguished from all other content and should not perpetuate gender or racial stereotypes.”²⁵

Aligned to this, the Committee have recommended that there is a need for the code to ensure that the profiling or targeting of children for commercial purposes is prohibited including practices that ‘rely on neuromarketing, emotional analytics, immersive advertising and advertising in virtual and augmented reality environments to promote products, applications and services’.²⁶

- Australian research found that during each hour that a child spends on the internet on their mobile device, they would see more than 17 food and beverage promotions, equating to 168 promotions per week and 8736 promotions per year. For each hour increase in usual time on the internet on mobile devices per week, children’s exposure to food promotions was found to increase by 6%.²⁷
- Canadian research found that children likely see food marketing in social media apps 111 times per week on average. On a yearly basis, this means that children likely see an average of 5772 instances of food marketing per year in these applications. 97% of those were for products high in fat, sugar or salt.²⁸

²⁴ Council of Europe, Guidelines to respect, protect and fulfil the rights of the child in the digital environment (2018) Recommendation CM/Rec(2018)7 of the Committee of Ministers, 20

²⁵ UN Committee on the Rights of the Child, General Comment no 25 (2021) on children’s rights in relation to the digital environment, CRC/C/GC/25, para 41

²⁶ UN Committee on the Rights of the Child, General Comment no 25 (2021) on children’s rights in relation to the digital environment, CRC/C/GC/25, para 42

²⁷ Kelly B, Bosward R, Freeman B. Australian children’s exposure to, and engagement with, web-based marketing of food and drink brands: cross-sectional observational study. *J Med Internet Res* 2021;23:e28144. doi:10.2196/28144pmid:http://www.ncbi.nlm.nih.gov/pubmed/34255675

²⁸ Harris JL, Kalnova SS. Food and beverage TV advertising to young children: Measuring exposure and potential impact. *Appetite*. 2018 Apr 1;123:49-55. doi: 10.1016/j.appet.2017.11.110. Epub 2017 Dec 5. PMID: 29217390.

- Further Canadian research estimated that children and adolescents see food marketing 30 and 189 times on average per week on social media apps, respectively.²⁹
- Nearly one in five 11–19-year-olds in the UK recall seeing junk food adverts on social media every day and two thirds remember seeing them at least weekly³⁰

The Code should look to ensure that a consistent feature for VSPS providers is introduced across all platforms that places a stringent requirement on users to declare when videos contain advertising and/or commercial communications. It should include a specific requirement for what form the declaration should take. This should be clear, concise, transparent and easy for children and young people to understand. ³¹

29 Potvin Kent M, Pauzé E, Roy E, de Billy N, Czoli C. Children and adolescents' exposure to food and beverage marketing in social media apps. *Pediatr Obes* 2019 Jun;14(6):e12508

³⁰ Critchlow N. et al (2020). Awareness of marketing for high fat, salt or sugar foods, and the association with higher weekly consumption among adolescents: A rejoinder to the UK government's consultations on marketing regulation. *Public Health Nutrition*, 23(14), 2637-2646

³¹ Children's Rights Alliance Submission to Coimisiún na Meán Call For Inputs on Developing First Online Safety Code 2023

Question 9: How should we ask VSPS providers to introduce and design a flagging mechanism in the Code? How can we ensure that VSPS providers introduce the mechanism in a user-friendly and transparent way? How should we ask VSP Providers to report the decisions they've made on content after it has been flagged?

It should not be expected or assumed that a child will be able to identify or report content or conduct which are against a service's community guidelines. They may not know if they themselves have breached a service's terms or what to do when something goes wrong, or how a service will respond when they have a problem that needs attention. They may be hesitant to report problems if they are concerned that they will get into trouble. While swift, effective reporting is an important provision for children, it is not the 'central' mechanism for protecting users. Relying on user reporting requires a child to understand the harm and their rights to be treated differently. This is simply not the reality for many children. It is welcome that the guidance encourages providers with a high number of users under the age of 18 to consider the needs of this group when designing or reviewing reporting/flagging systems (4.60), but this should be a requirement, not something providers only need to 'consider'.³²

The best interest of the child should be a key focus when considering the design of the flagging mechanism in the code. The Council of Europe (COE) Guidelines to Respect, Protect and Fulfil the Rights of the Child in the Digital Environment provide that 'in all actions concerning children in the digital environment, the best interests of the child shall be a primary consideration' and further recommend that States should strike a balance between the child's right to protection and their other rights to freedom of expression, participation and access to information. The COE also acknowledges the differing levels of maturity and understanding of children at different ages and recommends that States recognise the evolving capacities of children which can mean that the 'policies adopted to fulfil the rights of adolescents may differ significantly from those adopted for younger children'.³³

³² Children's Rights Alliance Submission to Coimisiún na Meán Call For Inputs on Developing First Online Safety Code 2023

³³ Council of Europe, 'Recommendation CM/Rec(2018)7 of the Committee of Ministers to Member States on Guidelines to respect, protect and fulfil the rights of the child in the digital environment' (COE 2018) Taken from Children's Rights Alliance Submission to Coimisiún na Meán Call For Inputs on Developing First Online Safety Code 2023

Question 10: What requirements should the Code include about age verification and age assurance? What sort of content should be shown by default to users who are logged out or in private browsing mode and whose age cannot be verified or assured? What evidence is there about the effectiveness of age estimation techniques? What current practices do you regard as best practice? Where accounts are not age verified should default privacy settings be used, should content default to universal content and should contact by others be more limited?

Question 21: Do you have any views on how requirements for commercial content arranged by a VSPS provider itself should be reflected in the Code?

Recommendations:

The Online Safety Code must ensure that children are protected effectively from harmful marketing and that their Convention on the Rights of the Child rights are upheld. This includes addressing commercial communications for mixed audiences, in order to capture all the marketing that children are exposed to.

Online Safety Codes should protect all children, not just those old enough to have digital access.

The online advertising ecosystem is complicated and includes services within, and also beyond, the scope of the Online Safety and Media Regulation Act.

Particular challenges arise in defining advertising to children and this has become a pertinent issue online also, as the Internet locations most visited by children are often not those “directed at” or “targeting” them but those providing access to a wide range of content (e.g. Google, Facebook, Instagram, YouTube).

Therefore, in determining the scope of harmful content and commercial communications, the Online Safety Code must ensure that children are protected effectively from harmful marketing and that their Convention on the Rights of the Child rights are upheld. This includes addressing commercial communications for mixed audiences, in order to capture all the marketing that children are exposed to.

When examining the features of online safety and media regulation as they relate to digital media, it is important to note that the potential for persuasive impact to be amplified is considerable, and so is the potential for the exploitation of child-consumers, as marketers take advantage of structural features of digital media platforms, particularly those offered by social media. This results often in the blurring of boundaries between marketing and media content.³⁴

Leading experts and academics have called for a cautious, expansive interpretation of children’s potential exposure to online commercial advertising or marketing. This is because

³⁴ Tatlow-Golden, M. and Garde, A. (2020). Digital food marketing to children: Exploitation, surveillance and rights violations. *Global Food Security* 27 (2020) 100423 <https://www.sciencedirect.com/science/article/pii/S2211912420300778> [Open Access]

companies in the digital ecosystem operate behind ‘walled gardens’, in the absence of transparency regarding online marketing strategies. They argue that such an approach is necessary to meet the dual objectives of protecting children from actual exposure to harmful marketing without restricting their right to participation.

Moreover, when considering harmful communications that impinge on the rights of children, commercial communications to or at children alone, should not just be considered. While “women are the primary targets of formula milk marketing and have been for decades... Approaches aim to engage women early in their pregnancies to create brand loyalty from then through their children’s infancy, the toddler years and beyond”³⁵ and these advertising strategies directly undermine children’s health and development. Online Safety Codes should protect all children, not just those old enough to have digital access. Babies and infants are our most vulnerable children and their protection should be extended through the caregiver by shielding the caregiver from infant formula marketing messages. The UN Convention on the Rights of the Child identifies implementation of the International Code of Marketing of Breast-milk Substitutes and strengthening the State’s regulatory framework for industries and enterprises to ensure that their activities do not have adverse impacts on children’s rights as crucial steps to upholding the Convention on the Rights of the Child.

³⁵ How the marketing of formula milk influences our decisions on infant feeding. Geneva: World Health Organization and the United Nations Children’s Fund (UNICEF), 2022. Licence: CC BY-NC-SA 3.0 IGO [Online] Available from: <https://www.unicef.org/eca/media/20086/file/Multi-country%20study%20examining%20the%20impact%20of%20BMS%20marketing%20on%20infant%20feeding%20decisions%20and%20practices.pdf> p14

Question 16: What requirements should the Code include about procedures for complaint-handling and resolution, including out-of-court redress or alternative-dispute resolution processes? To what extent should these requirements align with similar requirements in the DSA? What current practices could be regarded as best practice? How frequently should VSPS providers be obliged to report to the Commission on their complaint handling systems and what should those reports contain? Should there be a maximum time-period for VSPS providers to handle user complaints and if so, what should that period be?

Recommendations:

The Commission should be able to assess the effectiveness of procedural measures against a set of statutory objectives that go beyond simplistic content-related benchmarks such as removal rates and response times.

The Commission should have the power to demand any type of granular information that is necessary for it to fulfil its supervisory tasks. Shifting scrutiny towards these processes would help address some of the causal factors that give rise to harmful content online.

Strong, proactive enforcement mechanisms are needed, which would apply stronger punitive measures for instances of noncompliance.

Reporting Requirements

The importance of transparency on the part of the services and platforms being regulated, and of the regulatory rules that are imposed on them, must be paramount. In the first instance, platforms and on-demand providers must respond to requests for information from the Commission. Currently, information in the public domain about platforms' approaches to dealing with harmful content is limited, with inconsistencies in the information that is available across platforms - there is no way of assessing the impact and effectiveness of these approaches, either with respect to takedown of material or blocking of legal content. Evaluations are generally conducted by intermediaries and platforms themselves, who have discretion on what to measure and disclose, with the transparency reports provided by many platforms noted not to "represent a comprehensive assessment of the impact of their content governance activities."

Indeed, it has been noted that outside of proprietary industry research, there is no independent public data to reliably monitor the extent to which children are exposed to commercial advertising and marketing online, and the impact these powerful and opaque digital marketing strategies have on children's identities, behaviour and development.³⁶

Complaint Handling and Self-Regulation

Currently, the Advertising Standards Authority of Ireland (ASAI) code regulates advertising – including online advertising – in Ireland, but this code is self-regulatory. The Irish Heart Foundation believes that the sections of the ASAI Code relating to the marketing of food and

³⁶ Garde, A et al. (2020). General Comment submission Children's rights in relation to the digital environment. [Online]. Available from: <https://www.ohchr.org/EN/HRBodies/CRC/Pages/GCChildrensRightsRelationDigitalEnvironment.aspx>

beverages to children are weak. In contrast to the statutory rules for television advertising which require advance clearance of ads, the ASAI only investigates complaints made about potential breaches of the Code after the ad in question has been seen by the public.

Problems with self-regulatory complaints mechanisms include:

- Complaint procedures do not provide a level playing field between citizens and industry: they are onerous and time-consuming processes for individual complainants.
- There is a lack of effective enforcement mechanisms such as fines to serve as a deterrent.
- Compliance and informal resolution processes are not open to public scrutiny.

The current enforcement mechanisms in place for non-broadcast commercial communications - of breaches being resolved by responding to individual complaints and promoting voluntary cooperation with the restriction – amounts to self-regulation, which has been shown to be ineffective³⁷³⁸³⁹ and thus will not achieve the aim to minimise the harms associated with children’s exposure to commercial communications.

The failures of self-regulation as well as the recommendations that the Media Commission will not co-operate with self-regulatory systems in the regulation of commercial communications and that non-statutory mechanisms are not considered as part of the regulatory framework, are considered in greater detail in the response to Question 19.

³⁷ World Cancer Research Fund International (2020). Building Momentum: lessons on implementing robust restrictions of food and non-alcoholic beverage marketing to children. Available at wcrf.org/buildingmomentum

³⁸ Boyland, E.J. and Harris, J.L., (2017). Regulation of food marketing to children: are statutory or industry self-governed systems effective?. *Public Health Nutrition*, 20(5), pp.761- 764.

³⁹ Reeve, B. and Magnusson, R., (2018). Regulation of food advertising to children in six jurisdictions: a framework for analyzing and improving the performance of regulatory instruments. *Ariz. J. Int'l & Comp. L.*, 35, p.71

Question 18: What approach do you think the Code should take to risk assessments and safety by design? Are there any examples you can point us towards which you consider to be best practice?

Recommendation:

Child rights impact assessment (CRIA) should be mandated

UNICEF and the WHO have recommended that in order to ensure that children's best interests are adequately considered in food marketing restrictions, governments should consider carrying out an ex-ante child rights impact assessment (CRIA).⁴⁰ CRIAs should help ensure that the best interests of children are taken into consideration during the policy and legislation development process and what the impact will be. Indeed, "ensuring that the best interests of the child are a primary consideration in business related legislation and policy development and delivery at all levels of government demands continuous child-rights impact assessments"⁴¹ and, as such, should not be overlooked in the development of Online Safety Codes.

⁴⁰ UNICEF and WHO. (2023). Taking action to protect children from the harmful impact of food marketing: a child rights-based approach. Geneva: World Health Organization and the United Nations Children's Fund (UNICEF). [Online] Available from: [https://www.unicef.org/media/142621/file/UNICEF-](https://www.unicef.org/media/142621/file/UNICEF-WHO%20Toolkit%20to%20Protect%20Children%20from%20the%20Harmful%20Impact%20of%20Food%20Marketing.pdf)

[WHO%20Toolkit%20to%20Protect%20Children%20from%20the%20Harmful%20Impact%20of%20Food%20Marketing.pdf](https://www.unicef.org/media/142621/file/UNICEF-WHO%20Toolkit%20to%20Protect%20Children%20from%20the%20Harmful%20Impact%20of%20Food%20Marketing.pdf)

⁴¹ UN Committee on the Rights of the Child (CRC), General comment No. 16 (2013) on State obligations regarding the impact of the business sector on children's rights, 17 April 2013, CRC/C/GC/16, available at: <https://www.refworld.org/docid/51ef9cd24.html> [accessed 14 August 2023] paragraph 78

Question 19: How do you think that cooperation with other regulators and bodies can help us to implement the Code for VSPS?

Recommendations:

A dedicated function within the Media Commission should relate to online harms as they relate to data protection. As recommended by the Data Protection Commission, online harms that relate to data protection should be dealt with by the Media Commission.

Self-regulatory bodies should not be involved in the regulation of commercial communications or in the implementation of the Online Safety Code for VSPs

Priorities:

- The objectives of addressing online harms on VSPs cannot be met in isolation without deep engagement with other regulators and consideration of interrelated issues, such as Data Protection, with the Data Protection Commissioner. The Online Safety Code should emphasise the extent to which online safety issues are interconnected with complex issues of data protection and privacy.
- Voluntary codes of practice should not be considered as a legitimate mechanism within the regulatory framework for online safety and should not be relied upon to stop harmful content online. Statutory mechanisms should be the sole structures by which Online Safety Codes are designed, implemented and enforced.

Data Protection

In the Submission by the Data Protection Commission to the Joint Committee on Media, Tourism, Arts, Culture, Sport and the Gaeltacht on the General Scheme of the Online Safety and Media Regulation Bill, the DPC referenced a regulatory lacuna and noted:

“In order to harness the full benefits of an Online Safety Commissioner as a constituent of the Media Commission and achieve meaningful outcomes for the public in this heretofore unregulated area, the DPC respectfully suggests that the Committee give due consideration to the following issues... Where a complaint or concern is raised about online content due to the harmful effects that content may have on the health/ safety/ wellbeing of one or more individuals, it should be dealt with through the regulatory framework envisaged by the OSMRB and via the enforcement powers of the Media Commission (i.e. acting through an Online Safety Commissioner). It is possible that a single piece of content may be considered as falling within multiple categories of harmful online content, and the DPC believes that the possibility of such material also infringing multiple areas of law (including data protection) should be addressed within the OSMRB. Specifically, the DPC is strongly of the view that “material that violates [data protection or privacy law]” should **absolutely not** be excluded from the scope of harmful online content in Part 4.”⁴²

⁴² Data Protection Commission. (2021). Submission by the Data Protection Commission to the Joint Committee on Media, Tourism, Arts, Culture, Sport and the Gaeltacht on the General Scheme of the Online Safety and Media Regulation Bill. [Online] Available from:

The DPC stressed that it was important that the Media Commission has the power to regulate all types of harmful online content, irrespective of whether they involve personal data. This is because there are clear limitations to the reach of data protection regulation, meaning it does not and cannot provide a comprehensive regime for tackling harmful content posted or shared in an online context.

While recognition is given in the Online Safety and Media Regulation Act legislation to harmonise some aspects of regulation of online safety that applies to data protection, the Online Safety Codes would benefit from a much more comprehensive understanding of the online harms related to data protection breaches through detailed explanation of how data protection online harms are to be addressed.

Indeed, while the Act provides that the Media Commission shall enter into memoranda of understanding with other relevant bodies, including the Data Protection Commission, there have been many criticisms levelled against the DPC on the capacity to fully and effectively execute its functions under the General Data Protection Regulation, with specific reference to its role as the Lead European Supervisory Authority in relation to large technology companies whose regional headquarters are located in Ireland.⁴³ This has meant that the data rights of citizens of the European Union are being threatened.

While there is no intention for the Media Commission to supplant the role of the DPC in relation to data protection and privacy matters in any way, there must be a dedicated resource within the Commission that can be seconded to work on online harms as they relate to data protection. This is of particular importance given, as the Call for Inputs notes, “some of Europe’s largest VSPS providers are based in Ireland and they provide large quantities of content to users in different languages and locations across the continent.” Such overlap between the activities of the Media Commission and the DPC or potential synergies are already set to be addressed through a memorandum of understanding, which can be updated as needs be, but this additional resource can ensure that the burden of online harms pertaining to data protection is sufficiently addressed, especially if the DPC is overstretched.

Self and Co-Regulation

“To prevent harm to people’s health and fulfil their obligation under the right to health, States should put in place national policies to regulate advertising of unhealthy foods. States should formulate laws and a regulatory framework with the objective of reducing children’s exposure to powerful food and drink marketing... Companies often voluntarily adopt self-formulated guidelines and standards to restrict Government regulation and respond public demands... However, self-regulation by companies has not had any significant effect on altering food marketing strategies... Due to a variety of reasons, such as the non-binding nature of such self-regulation, lack of benchmarks and transparency,

https://data.oireachtas.ie/ie/oireachtas/committee/dail/33/joint_committee_on_tourism_culture_arts_sport_and_media/submissions/2021/2021-11-02_submission-data-protection-commission_en.pdf

⁴³ ICCL. (2021). ICCL alerts Irish Government of strategic economic risk from failure to uphold the GDPR. [Online] Available from: <https://www.iccl.ie/news/iccl-alerts-irish-government-of-strategic-economic-risk-from-failure-to-uphold-the-gdpr/>

inconsistent definition of children and different nutrition criteria, companies may be able to circumvent guidelines, blunting the intended effect of marketing guidelines they instituted... Owing to the inherent problems associated with self-regulation and public-private partnerships, there is a need for States to adopt laws that prevent companies from using insidious marketing strategies.”⁴⁴

A 2013 systematic review⁴⁵ found significant divergence between the reported impact of marketing regulation (including self-regulation by industry) provided in peer-reviewed journals, or industry-sponsored reports, showing the need for external monitoring. Moreover, of studies evaluating voluntary policies, significantly more studies showed undesirable effects than desirable effects on exposure to, and power of, food marketing. This was not the case for studies evaluating mandatory policies.⁴⁶

Where those with commercial interests are involved in the development and wording of self-regulatory codes, the resulting provisions are often so weak or unclear that they are meaningless. The "commitments" they contain, for instance, are often expressed as weak targets or goals, with thresholds so low that companies can reach them without much effort, and they routinely include imprecise wording which is open to interpretation.

A 2023 report on protecting children from the harmful impact of food marketing from the World Health Organization and the United Nations Children’s Fund note that “the main stakeholders responsible for implementing effective policies to protect children from the harmful impact of food marketing should be trusted public authorities, as the bearers of a duty to protect children’s rights and public health. Delegation of responsibility to other stakeholders (e.g. sector associations representing the advertising industry or broadcasters) is not recommended as it has been shown to create conflicts of interest at the heart of policy discussions in many countries”.⁴⁷ Voluntary actions, such as industry-led pledges and other self-regulatory measures, have not been demonstrated to work effectively to protect children from the impact of food marketing and commercial communications. They are not – and should not be viewed as – an appropriate mechanism to ensure that children are effectively protected from harmful marketing. Furthermore, a child rights-based approach to the regulation of food marketing requires that competent public authorities do not engage in ineffective public-private partnerships amounting to the delegation of the mandate they have to protect child health and child rights to private business operators. and should therefore not be included as part of the regulatory package as part of the Online Safety Code.

⁴⁴ UN General Assembly Human Rights Council. (2014). Report of the Special Rapporteur on the right of everyone to the enjoyment of the highest attainable standard of physical and mental health, Anand Grover: Unhealthy foods, non-communicable diseases and the right to health. [Online] Available from:

https://www.ohchr.org/sites/default/files/HRBodies/HRC/RegularSessions/Session26/Documents/A-HRC-26-31_en.doc

⁴⁵ Galbraith-Emami, S. and Lobstein, T. (2013) ‘The impact of initiatives to limit the advertising of food and beverage products to children: a systematic review’. *Obesity Reviews*.

⁴⁶ Boyland, E, McGale, L, Maden, M, Hounsome, J, Boland, A, Jones, A. Systematic review of the effect of policies to restrict the marketing of foods and non-alcoholic beverages to which children are exposed. *Obesity Reviews*. 2022; 23(8):e13447. doi:10.1111/obr.13447

⁴⁷ UNICEF and WHO. (2023). Taking action to protect children from the harmful impact of food marketing: a child rights-based approach. Geneva: World Health Organization and the United Nations Children’s Fund (UNICEF). [Online] Available from: [https://www.unicef.org/media/142621/file/UNICEF-](https://www.unicef.org/media/142621/file/UNICEF-WHO%20Toolkit%20to%20Protect%20Children%20from%20the%20Harmful%20Impact%20of%20Food%20Marketing.pdf)

[WHO%20Toolkit%20to%20Protect%20Children%20from%20the%20Harmful%20Impact%20of%20Food%20Marketing.pdf](https://www.unicef.org/media/142621/file/UNICEF-WHO%20Toolkit%20to%20Protect%20Children%20from%20the%20Harmful%20Impact%20of%20Food%20Marketing.pdf)
p26

Some research related to self-regulation:

Boyland, E.J. and Harris, J.L., (2017). Regulation of food marketing to children: are statutory or industry self-governed systems effective? *Public Health Nutrition*, 20(5), pp.761- 764.

Hawkes, C. (2008). Agro-food industry growth and obesity in China: what role for regulating food advertising and promotion and nutrition labelling?. *Obesity Reviews*, 9, 151-161.

Kunkel, D. L., Castonguay, J. S., & Filer, C. R. (2015). Evaluating industry self-regulation of food marketing to children. *American Journal of Preventive Medicine*, 49(2), 181-187.

León-Flández, K., Rico-Gómez, A., Moya-Geromin, M. Á., Romero-Fernández, M., Bosqued-Estefania, M. J., Damian, J., ... & Royo-Bordonada, M. A. (2017). Evaluation of compliance with the Spanish Code of self-regulation of food and drinks advertising directed at children under the age of 12 years in Spain, 2012. *Public Health*, 150, 121-129.

Mackay, S. (2009). Food advertising and obesity in Australia: to what extent can self-regulation protect the interests of children. *Monash UL Rev.*, 35, 118.

Reeve, B. and Magnusson, R., (2018). Regulation of food advertising to children in six jurisdictions: a framework for analyzing and improving the performance of regulatory instruments. *Ariz. J. Int'l & Comp. L.*, 35, p.71

Sing, F., Mackay, S., Culpin, A., Hughes, S., & Swinburn, B. (2020). Food advertising to children in New Zealand: A critical review of the performance of a self-regulatory complaints system using a public health law framework. *Nutrients*, 12(5), 1278.

Thornley, L., Signal, L., & Thomson, G. (2010). Does industry regulation of food advertising protect child rights?. *Critical Public Health*, 20(1), 25-33.

World Cancer Research Fund International (2020). Building Momentum: lessons on implementing robust restrictions of food and non-alcoholic beverage marketing to children. Available at wcrf.org/buildingmomentum

Question 22: What compliance monitoring and reporting arrangements should we include in the Code?

Recommendations:

On the issue of monitoring and enforcement, the Irish Heart Foundation endorses the processes and actions put forward by UNICEF and the WHO⁴⁸ in terms of protecting children from harmful food marketing:

- the application of deterrent sanctions for non-compliance. Enforcement mechanisms should be both reactive and proactive, meaning that they should be open to both receiving notification of infringements, and detecting infringements through screenings and ongoing monitoring.
- continuous monitoring and enforcement mechanisms should be established (including a complaints procedure available to those with a legitimate complaint)
- clear authority to enforce the restrictions

Furthermore:

- Regulated entities should not just be required to “provide periodic reports on their compliance or otherwise with codes”, but should also be forced to provide any type of granular information to the Commission that is necessary for it to fulfil its supervisory tasks
- Provision should be made to enable independent public interest research, based on data from platforms

Codes of conduct do not, by definition, include meaningful sanctions for those who do not comply with the code, or who are found to be in breach. Rather, industry use the threat of “reputational damage” as an adequate deterrent for companies from breaching these codes. Voluntary codes are particularly susceptible to breaches of all or some of their provisions when it is more commercially advantageous to do so and, in the absence of sanctions for non-compliance, companies will continue to flaunt the code. This is especially true if there isn’t a public awareness of the code or the complaints process. Appropriate sanctions must be set for non-compliance- It is not enough to rely on the censure of civil society and the media for failure to comply. Failure to comply with restrictions established through laws or regulations must lead to the application of effective sanctions.

⁴⁸ UNICEF and WHO. (2023). Taking action to protect children from the harmful impact of food marketing: a child rights-based approach. Geneva: World Health Organization and the United Nations Children’s Fund (UNICEF). [Online] Available from: <https://www.unicef.org/media/142621/file/UNICEF-WHO%20Toolkit%20to%20Protect%20Children%20from%20the%20Harmful%20Impact%20of%20Food%20Marketing.pdf>

Re: Call of Inputs: Developing Ireland's First Binding Online Safety Code for Video-Sharing Platform Services

Submission on behalf of the National Suicide Research Foundation (NSRF)

Introduction

The National Suicide Research Foundation welcome the call of inputs in relation to developing Ireland's First Binding Online Safety Code for Video-Sharing Platform and the proactive approach of the partner agencies in this regard, including the HSE National Office for Suicide Prevention, Samaritans Ireland, Headline and the HSE Department of Health.

Online Safety is one of the key priority areas identified in Ireland's National Strategy to Reduce Suicide 2015-2024, Connecting for Life, specifically Strategic Action 1.4.1 – 'engage with online platforms to encourage best practice in reporting around suicidal behaviour, so as to encourage a safer online environment in this area'. The National Suicide Research Foundation (NSRF) is a Connecting for Life funded agency and is recognised as a World Health Organisation (WHO) Collaborating Centre for Surveillance and Research in Suicide Prevention, where our remit has included developing a [WHO resource for filmmakers and others working on stage and screen on preventing suicide](#).

The need for improved online safety is also underlined by the United Nations' Sustainable Development Goals (SDGs), in particular Goal 3.4: By 2030, reduce by one third premature mortality from non-communicable diseases, including suicide, through prevention and treatment and promote mental health and wellbeing (UN, 2015).

International literature suggests that there is inadequate understanding of the different forms of self-harm and suicide online, including a lack of definition and taxonomy of self-harm and suicide content on social media (Scherr, 2022), with a paucity of content definitively classifiable as explicitly harmful or helpful (Brennan et al. 2022). Nevertheless, research indicates that the internet and social media are double edged swords and can provide both benefits and challenges (Robinson et al. 2016; Fu et al. 2013). A strong body of research suggests that there is significant risk of harm related to online behaviour such as reinforcement, stigmatization, normalisation, triggering and contagion, in addition to hindering professional help-seeking and the depiction of methods of suicidal behaviour (Marchant et al. 2017; Lewis & Seko 2016 ; Daine et al. 2013 ; Alao et al. 2006).

Question 1: What do you think our main priorities and objectives should be in the first binding Online Safety Code for VSPS? What are the main online harms you would like to see it address and why?

In November 2020, the National Suicide Research Foundation completed a literature review on the harmful impact of suicide and self-harm content online in collaboration with the National Office for Suicide Prevention. The review was further updated in August 2023. In total 150 peer reviewed publications were included and the following types of harmful content online related to suicidal behaviour and self-harm were identified, all of which should be considered priority areas for the first binding Online Safety Code for VSPS:

➤ **(1) Online information sources (websites and platforms used to inform method)**

- Easy access to information about suicide methods and pro-suicide web sites on the internet may influence a small but significant number of suicides (Gunnell et al. 2012).
- Particularly new or emerging methods of self-harm or suicide may be promoted online as well as facilitating access to these methods (Paul et al. 2017; Chang et al. 2015).
- There is a need for longer-term preventive action in relation to dissemination of suicide methods (Gunnell et al. 2012), and harmful information on the internet (see Q2 for recommendations).

➤ **(2) Search engines**

- Conflicting results make it difficult to draw definitive conclusions with regard to search engines. Studies have identified a positive association between suicide rates and search volume (Arendt, 2018; Chandler, 2018), including an analysis of data from fifty countries across five continents (Arendt, 2018).
- However, others have found no relationship between “suicide” or “suicide and methods” searches and suicide incidence (Waszak et al. 2018; Bruckner et al. 2014; Sueki et al. 2011).
- The implementation of evidence-informed guidelines for sites and platforms hosting user generated content is recommended (The Samaritans, 2020).

➤ **(3) Social networking sites**

- Harmful aspects of social networks include normalising self-harming behaviour; the inclusion of dialogue around motivation and triggers for self-harm, increased suicidal ideation or plans among users; and accessible depictions of self-harm acts (Dyson et al. 2016).
- For adolescents, greater time spent on social networks has been associated with increased self-harm behaviour and suicidal ideation, linked to users receiving damaging messages promoting self-harm, copying self-harming behaviour of others, and emulating self-harm practices from shared videos. Time expended on social networks has also been found to lead to amplified psychological distress, an unfulfilled need for mental health support, adverse self-rated mental health, and increased suicidal ideation (Muslic et al. 2023; Winstone et al. 2022; Gámez-Guadix et al. 2022; Sumner et al. 2021; Shafi et al. 2020 ; Memon et al. 2018 ; Berryman et al. 2017), particularly among girls (Luby & Kertz, 2019) and among sexual and gender minority youth (Nesi et al. 2021).

- Although suicide-related online experience is a common, and likely underestimated, precursor to suicide in young people, its contributing role remains unclear (Rodway et al 2022) but social media use may be an indicator of impulsivity (Shafi et al. 2021).
- Yet other research suggests that there are null, mixed or very small associations between time spent online and mental health problems for most adolescents (Orben & Przybylski, 2019 ; Best et al. 2014).
- A focus on safe browsing warrants consideration, in addition to tools that limit time and diversify content (Brennan et al. 2022).

Facilitate access to potentially harmful information

- Social networks may provide access to suicide content and violent images (Daine et al. 2013). Non-Suicidal Self Injury (NSSI) behaviours are becoming common across social networks, particularly Instagram (Brown et al.2018).

Online contagion

- Social media may facilitate contagion and clusters by spreading suicidal thoughts and acts, however it may also have a positive role in supporting people at risk for suicide (Kline et al 2023; Swedo et al. 2020 ; Brown et al. 2020 ; Fu et al. 2013)
- Memorial pages may also lead to social contagion and facilitate the rapid spread of information about deaths by suicide in the community (Robertson et al, 2012).

Normalisation of self-harming behaviour

- By viewing and reading material on social networks and pro-suicide websites a normalisation of self-harm may take place (Dyson et al. 2016; Daine et al. 2013; Lewis & Baker, 2011), which may perpetuate associated beliefs and behaviours and hinder access to treatment (Hilton, 2017).
- The sense of companionship and community generated on Twitter specifically may facilitate the normalisation of self-harm (Marchant et al. 2021; Hilton, 2017).
- Protection and safety frameworks, in addition to voluntary industry codes of conduct to prevent normalisation of harmful behaviour related to suicide and self-harm should be considered (The Lancet, 2019)

Celebrity suicide

- Suicide deaths of celebrities of high prominence, can lead to considerable national increases in internet search volumes for suicide-related terms (Ortiz et al. 2018) and the content of posts may show considerable changes that suggest increased suicidal ideation (Kumar et al. 2015).

- Deaths of younger celebrities may generate a higher number of posts (Ueda et al. 2017)
- It is important that suicide deaths are reported sensibly and responsibly in the media in compliance with the media guidelines for reporting suicide (Ortiz et al. 2018 ; Ueda et al. 2017 ; The Samaritans 2013). Platforms should be vigilant of harmful activity following deaths of celebrities of high prominence.

Cyberbullying

- Cyberbullying is a risk factor for self-harm and suicide in patients with mental health problems (Hellstand et al 2021).
- Cyber-only bullying appears to be related to specific mental health issues beyond those associated with school-only bullying (Ossa et al.2023)
- Victims of cyberbullying are at a greater risk of both self-harm and suicidal behaviours than non-victims (John et al. 2018).

Suicide notes

- Evidence of copycat suicides induced by suicide notes on social networking sites is unclear. However, notes may facilitate immediate intervention from other users (Ruder et al. 2011).

➤ (4) Online imagery and videos

- Viewing self-harm images online may have both harmful and protective effects, but harmful effects are more prevalent (Susi et al. 2023).
- There has been an increase in harmful graphic self-harm imagery over time with an absence of moderation, anonymity, and pictures easy to find using the search function (Marchant et al. 2021).
- Depictions of self-harm acts through imagery and video may empower the normalisation of young people's self-harm and pictures may incite a physical reaction and stimulate behavioural enactment (Jacob et al. 2017).
- Graphic content aligned to self-harm is prevalent on Instagram in particular (Miguel et al. 2017) and is often obscured by unclear or secret hashtags, while subliminal messages are ethically highly problematic (Arendt et al 2021). Time scrolling on Instagram has also been associated with normalization of self-harm and contagion (Moss et al. 2022), while content advisory warnings on this platform may not be dependable (Moreno et al. 2016).
- Videos related to self-harm are common on YouTube and may encourage the normalisation of self-harm and may enhance the behaviour through regular viewing of graphic videos (Lewis et al. 2011). However, YouTube may also provide an

opportunity to engage with teenagers and to promote positive mental health (Dagar & Falcone, 2020; Lewis et al. 2018).

- Portrayals of fictional suicides in films/ tv series have been associated with a significant increase in suicide rates (Bridge et al. 2020) and an increase in self-harm admissions (Cooper et al 2018) , among young people.
- There is a need for longer-term preventive action in relation to dissemination of images related to self-harm (Brown et al. 2018) and harmful information on the internet (see Q2 for recommendations).

➤ (5) Online forums or message boards

- Forums or message boards may normalise and promote self-injurious and suicidal behaviour and expose new potentially lethal behaviours to those with a history of self-harm and those exploring identity options (Whitlock et al. 2006 ; Becker & Schmidt, 2004; Becker et al. 2004).
- Yet online communications can offer important social support for otherwise isolated adolescents (Whitlock et al. 2006).
- A focus on safe browsing warrants consideration, in addition to tools that limit time and diversify content (Brennan et al. 2022).

➤ (6) Pro-suicide and self-harm websites

- Adverse effects of visiting pro-suicide and self-harm websites include victimization, exacerbated self-harm, triggering of behaviour, seeking a partner to take your own life with and searching for highly lethal methods (Mokkenstorm et al. 2020; Minkkinen et al. 2017; Harris & Roberts, 2013).
- Young people who visited such websites were seven times more likely to say they had thought about killing themselves and 11 times more likely to think about hurting themselves even after adjusting for several known risk factors for thoughts of self-harm and thoughts of suicide (Mitchell et al. 2014).
- Self-harm forums and internet message boards, alternatively, may have a constructive effect on most users (Eichenberg & Schott, 2017). Suicide-related internet use is multifaceted, and impact cannot be directly attributed to explicit types of websites or online content (Mok et al. 2015).
- Professionals and stakeholders working in the area of suicide prevention need to be mindful of the existence and potential risk of such websites and communicate with youth in a meaningful, balanced way about them to promote safety and indicate risk (Biddle et al. 2018 ; Mitchell et al. 2014).

- In 2006, Australia became the first country to prohibit pro suicide and self-harm websites (Pirkis et al. 2009). New Zealand and the United Kingdom have followed suit in either passing or recently amending legislation to hold individuals who assist, encourage, aid, provide guidance or procure a suicide or suicide attempt online accountable (Phillips et al. 2019 ; Cheng 2011).

➤ (7) Online suicide 'games'

- Pro-suicide games or messages online such as the 'Blue Whale Challenge' can circulate quickly and globally (Sumner et al. 2019), particularly among vulnerable adolescents (Lupariello et al. 2019).
- The Blue Whale Challenge illustrates how social media can glorify, normalize and reinforce self-harming and suicidal behaviours (Khasawneh et al. 2020), and amplify suicide contagion among vulnerable cohorts (Upadhyaya & Kozman 2022).
- Early detection of coercions related to suicide and mental health is required and further research to identify emerging harms in real time is needed (Sumner et al. 2019). Safeguards must be introduced to stop content from being posted and children and adolescents viewing it (Upadhyaya & Kozman 2022) (see Q2 for recommendations).

➤ (8) The 'Darknet'

- People may be more exposed to harmful suicide and self-harm content when using the 'darknet'. Many darknet search engines facilitate access to forums that are pro-suicide and blocked or filtered by most of the surface web search engines (e.g. Google) (Morch et al. 2018).
- There is a need to consider the so called 'Darknet' when developing codes

➤ (9) Livestream suicide/Cybersuicide

- The possibility of livestream suicides initiating a suicide contagion (the Werther Effect) has been identified as a concern in the literature (Birbal et al. 2009), however, there is a lack of research on this area.

➤ (10) Online Suicide 'pacts'

- Suicide 'pacts' are an agreed plan between two or more individuals to take their own life. Twitter, specifically, has been identified as a potential attractive place where people try to meet others to make a suicide pact. This may be due to particular features such as a user's ability to create several accounts with different names but without disclosing much personal information (Lee & Kwon, 2018), which poses challenges to intervene in a timely manner.

Question 2: What types of online harms do you think should attract the most stringent risk mitigation measures by VSPS? How could we evaluate the impact of different types of harms e.g. severity, speed at which harm may be caused? Is there a way of classifying harmful content that you consider it would be useful for us to use?

➤ ***Most stringent measures***

(1) Reducing the dissemination of methods and harmful imagery

There is an urgent need for longer-term preventive action in relation to dissemination of suicide methods (Gunnell et al. 2012), images related to self-harm (Brown et al. 2018) and harmful information on the internet.

Internet service providers should be encouraged to regularly review content and advisory notices (Moreno et al. 2016), remove pro-suicide sites promoting the use of high-lethality methods (Gunnell et al. 2015) and take appropriate measures for preventing online social contagion (Brown et al. 2018). Monitoring and regulating online information on methods may also be beneficial (The Samaritans, 2020 ; Chang et al. 2015).

The implementation of evidence-informed guidelines for sites and platforms hosting user generated content is recommended (The Samaritans, 2020). However, mechanisms to limit or prohibit harmful content must be implemented with caution to avoid causing unintentional harm (Lavis & Winter, 2020).

(2) Preventing normalisation of self-harm and suicidal behaviour

Social media and the accessibility of celebrity discourse can contribute to normalising self-harm which may prolong and exacerbate associated behaviours and delay help seeking (Hilton, 2017). Further research examining if social media facilitates or deters suicidal behaviour is warranted. Protection and safety frameworks, in addition to voluntary industry codes of conduct to prevent normalisation of harmful behaviour related to suicide and self-harm should be considered (The Lancet, 2019). A focus on safe browsing also warrants consideration, in addition to tools that limit time and diversify content (Brennan et al. 2022).

(3) Early detection of online ‘suicide games’

Novel online risks to mental health, such as pro-suicide games can circulate quickly and globally. Early detection of coercions related to suicide and mental health is required and further research to identify emerging harms in real time is needed (Sumner et al. 2019). Safeguards must be introduced to stop content from being posted and children and adolescents viewing it (Upadhyaya & Kozman, 2022). Enhanced attention on innovative approaches to identify threats may play an important role (Sumner et al. 2019). Empowering young people to share user experiences and contribute to online safety initiatives may produce positive outcomes (Biddle et al. 2022; Marchant et al 2021).

(4) Prohibit pro suicide and self-harm websites

Adverse effects of visiting pro-suicide and self-harm websites include victimization, exacerbated self-harm, triggering of behaviour, seeking a partner to take your own life with and searching for highly lethal methods (Mokkenstorm et al. 2020; Minkkinen et al. 2017; Harris & Roberts, 2013).

Self-harm forums and internet message boards, alternatively, may have a constructive effect on most users (Eichenberg & Schott, 2017). Suicide-related internet use is multifaceted, and impact cannot be directly attributed to explicit types of websites or online content (Mok et al. 2015).

In 2006, Australia became the first country to prohibit pro suicide and self-harm websites (Pirkis et al. 2009). New Zealand and the United Kingdom have followed suit in either passing or recently amending legislation to hold individuals who assist, encourage, aid, provide guidance or procure a suicide or suicide attempt online accountable (Phillips et al. 2019 ; Cheng 2011). In the US, the Children’s Internet Protection Act (CIPA) stipulates that schools and libraries must block access to harmful content online (Phillips et al. 2019). There have been calls for others to follow suit, however it remains a complex issue. The NSRF would advocate for increased collaboration across jurisdictions to achieve consistency and reduce access to these sites.

➤ ***Evaluating the impact of different types of harms***

Literature suggests that there is inadequate understanding of the different forms of self-harm and suicide online, including a lack of definition and taxonomy of self-harm and suicide content on social media (Scherr 2022), with a paucity of content definitively classifiable as explicitly harmful or helpful (Brennan et al 2022). Scheuerman et al’s 2021 framework of severity for harmful content online identify eight factors to measure severity (perspectives, intent, agency, experience, scale, urgency, vulnerability, sphere).

➤ ***Classifying harmful content***

The NSRF would recommend categorising the following types of online content in line with international research (McTernan & Ryan, 2023; Susi et al 2023, Marchant et al. 2017)

- 1) Online information sources (websites used to inform method)
- 2) Search engines
- 3) Social networks
 - Facilitate access to potentially harmful information
 - Facilitate contagion
 - Normalising self-harm and suicide
 - Increased risk following celebrity suicide
 - Facilitate cyberbullying
 - Suicide notes
- 4) Online imagery and videos
- 5) Online forums/message boards
- 6) Pro-suicide and self-harm sites
- 7) Online suicide ‘games’
- 8) The ‘Darknet’
- 9) Livestream suicide / cybersuicide
- 10) Online suicide ‘pacts’

Question 3: Do you have reports, academic studies or other relevant independent research that would support your views? If you do, please share them with us with links to relevant reports, studies or research.

Irish research

- McTernan N, Ryan F (2023). [The Harmful Impact of Suicide and Self-Harm Content Online: A Review of the Literature](#). *National Suicide Research Foundation*
- Benson R, McTernan N, Ryan F, Arensman E. [Suicide clustering and contagion: The role of the media](#). *Suicidologi*. 26(2)

International Systematic Reviews

- Susi K, Glover-Ford F, Stewart A, Knowles Bevis R, Hawton K. [Research Review: Viewing self-harm images on the internet and social media platforms: systematic review of the impact and associated psychological mechanisms](#). *J Child Psychol Psychiatry*. 2023 Aug;64(8):1115-1139
- Brennan, C , Saraiva, S, Mitchell, E et al. (2022) [Self-harm and suicidal content online, harmful or helpful? A systematic review of the recent evidence](#). *Journal of Public Mental Health*, 21 (1). pp. 57-69. ISSN 1746-5729
- Marchant A, Hawton K, Burns L, Stewart A, John A. [Impact of Web-Based Sharing and Viewing of Self-Harm-Related Videos and Photographs on Young People: Systematic Review](#). *J Med Internet Res*. 2021 Mar 19;23(3):e18048
- Marchant A, Hawton K, Stewart A, Montgomery P, Singaravelu V, Lloyd K, Purdy N, Daine K, John A. [A systematic review of the relationship between internet use, self-harm and suicidal behaviour in young people: The good, the bad and the unknown](#). *PLoS One*. 2018 Mar 1;13(3):e0193937.

Conclusion

This research area is rapidly evolving with a significant increase in the number of publications in recent years (Krysinka et al. 2017). It is clear, that as the ‘internet-native’ generation matures, suicide and self-harm related internet use is likely to become increasingly relevant and may be a proxy indicator for intent (Padmanathan et al. 2018). During the COVID-19 pandemic, people in Ireland and throughout the world spent more time online (CSO, 2020), and social media use was associated with some adverse mental health conditions, suicidal ideation, increased fear and anxiety (Draženiović et al 2023 ; Memon et al 2021). The NSRF would be happy to support the Coimisiún na Meán in developing codes and protocols in accordance with international best practice, as outlined above and in the recently updated literature review.

National Suicide Research Foundation

September 4th, 2023

References

- Alao AO, Soderberg M, Pohl EL, Alao AL. Cybersuicide: review of the role of the internet on suicide. *Cyberpsychol Behav*. 2006 Aug;9(4):489-93.
- Arendt F. Suicide rates and information seeking via search engines: A cross-national correlational approach. *Death Stud*. 2018;42(8):508-512
- Arendt F, Markiewitz A, Scherr S. Investigating Suicide-Related Subliminal Messages on Instagram. *Crisis*. 2021 Jul;42(4):263-269.
- Berryman C, Ferguson CJ, Negy C. Social Media Use and Mental Health among Young Adults. *Psychiatr Q*. 2018;89(2):307-314. doi:10.1007/s11126-017-9535-6
- Best P, Manktelow R, Taylor B. Online communication, social media and adolescent wellbeing. A systematic narrative review. *Child Youth Serv Rev* 2014; 41:27-36
- Biddle L, Derges J, Goldsmith C, Donovan JL, Gunnell D. Using the internet for suicide-related purposes: Contrasting findings from young people in the community and self-harm patients admitted to hospital. *PLoS One*. 2018 May 24;13(5):e0197712.
- Biddle L, Rifkin-Zybutz R, Derges J, Turner N, Bould H, Sedgewick F, Gooberman-Hill R, Moran P, Linton MJ. Developing good practice indicators to assist mental health practitioners to converse with young people about their online activities and impact on mental health: a two-panel mixed-methods Delphi study. *BMC Psychiatry*. 2022 Jul 19;22(1):485
- Birbal R, Maharajh HD, Birbal R, et al. Cybersuicide and the adolescent population: challenges of the future? *Int J Adolesc Med Health* 2009; 21:151–9
- Brennan, C , Saraiva, S, Mitchell, E et al. (4 more authors) (2022) Self-harm and suicidal content online, harmful or helpful? A systematic review of the recent evidence. *Journal of Public Mental Health*, 21 (1). pp. 57-69. ISSN 1746-5729
- Bridge JA, Greenhouse JB, Ruch D, Stevens J, Ackerman J, Sheftall AH, Horowitz LM, Kelleher KJ, Campo JV. Association Between the Release of Netflix's 13 Reasons Why and Suicide Rates in the United States: An Interrupted Time Series Analysis. *J Am Acad Child Adolesc Psychiatry*. 2020 Feb;59(2):236-243.
- Brown RC, Fischer T, Goldwich AD, Keller F, Young R, Plener PL. #cutting: Non-suicidal self-injury (NSSI) on Instagram. *Psychol Med*. 2018 Jan;48(2):337-346
- Brown RC, Fischer T, Goldwich DA and Plener PL. “I just finally wanted to belong somewhere” — Qualitative Analysis of Experiences With Posting Pictures of Self-Injury on Instagram. *Frontiers in Psychiatry*. 2020. 11:274. doi: 10.3389/fpsy.2020.00274
- Bruckner TA, McClure C, Kim Y. Google searches for suicide and risk of suicide. *Psychiatr Serv*. 2014 Feb 1;65(2):271-2.
- Chandler V. Google and suicides: what can we learn about the use of internet to prevent suicides? *Public Health*. 2018;154:144–50.

- Chang SS, Kwok SS, Cheng Q, Yip PS, Chen YY. The association of trends in charcoal-burning suicide with Google search and newspaper reporting in Taiwan: a time series analysis. *Soc Psychiatry Psychiatr Epidemiol*. 2015;50(9):1451-1461.
- Cheng Q. Are internet service providers responsible for online suicide pacts? *BMJ*. 2011 Apr 13;344:d2113.
- Cooper, M. T., Jr., Bard, D., Wallace, R., Gillaspay, S., & Deleon, S. (2018). Suicide attempt admissions from a single children's hospital before and after the introduction of Netflix series 13 Reasons Why . *Journal of Adolescent Health*, 63(6), 688–693.
- Dagar A, Falcone T. High Viewership of Videos About Teenage Suicide on YouTube. *J Am Acad Child Adolesc Psychiatry*. 2020 Jan;59(1):1-3.e1.
- Daine K, Hawton K, Singaravelu V, Stewart A, Simkin S, Montgomery P. The power of the web: a systematic review of studies of the influence of the internet on self-harm and suicide in young people. *PLoS One*. 2013 Oct 30;8(10):e77555
- Dyson MP, Hartling L, Shulhan J, Chisholm A, Milne A, Sundar P, Scott SD, Newton AS. A Systematic Review of Social Media Use to Discuss and View Deliberate Self-Harm Acts. *PLoS One*. 2016 May 18;11(5)
- Eichenberg C, Schott M. An Empirical Analysis of Internet Message Boards for Self-Harming Behavior. *Arch Suicide Res*. 2017 Oct-Dec;21(4):672-686.
- Fu KW, Cheng Q, Wong PW, Yip PS. Responses to a self-presented suicide attempt in social media: a social network analysis. *Crisis*. 2013 Jan 1;34(6):406-12.
- Gámez-Guadix M, Mateos E, Wachs S, Blanco M. Self-Harm on the Internet Among Adolescents: Prevalence and Association With Depression, Anxiety, Family Cohesion, and Social Resources. *Psicothema*. 2022 May;34(2):233-239
- Gunnell D, Bennewith O, Kapur N, Simkin S, Cooper J, Hawton K. The use of the Internet by people who die by suicide in England: a cross sectional study. *J Affect Disord*. 2012;141(2-3):480-483
- Harris IM, Roberts LM. Exploring the use and effects of deliberate self-harm websites: an Internet-based study. *J Med Internet Res*. 2013 Dec 20;15(12):e285.
- Hellstrand K, Rogers SC, DiVietro S, Clough M, Sturm J. Prevalence of Cyberbullying in Patients Presenting to the Pediatric Emergency Department. *Pediatr Emerg Care*. 2021 Jun 1;37(6):e334-e338
- Hilton EC. Unveiling self-harm behaviour: what can social media site Twitter tell us about self-harm? A qualitative exploration. *J Clin Nurs*. 2017 Jun;26(11-12):1690-1704.

- Jacob N, Evans R, Scourfield J. The influence of online images on self-harm: A qualitative study of young people aged 16-24. *J Adolesc.* 2017 Oct;60:140-147.
- John A, Glendenning AC, Marchant A, Montgomery P, Stewart A, Wood S, Lloyd K, Hawton K. Self-Harm, Suicidal Behaviours, and Cyberbullying in Children and Young People: Systematic Review. *J Med Internet Res* 2018;20(4):e129
- Khasawneh A, Chalil Madathil K, Dixon E, Wiśniewski P, Zinzow H, Roth R. Examining the Self-Harm and Suicide Contagion Effects of the Blue Whale Challenge on YouTube and Twitter: Qualitative Study. *JMIR Ment Health.* 2020 Jun 5;7(6):e15973.
- Kline M, Metcalf MM, Patel E, Chang EL, Nguyen MB. Adolescent Experiences With Social Media and Suicidality, *Academic Pediatrics*, Volume 23, Issue 4, 2023, Pages 755-761
- Kumar M, Dredze M, Coppersmith G, De Choudhury M. Detecting Changes in Suicide Content Manifested in Social Media Following Celebrity Suicides. *HT ACM Conf Hypertext Soc Media.* 2015 Sep;2015:85-94.
- Lavis, A., & Winter, R. #Online harms or benefits? An ethnographic analysis of the positives and negatives of peer-support around self-harm on social media. *Journal of Child Psychology and Psychiatry* 2020.
- Lee SY, Kwon Y. Twitter as a place where people meet to make suicide pacts. *Public Health.* 2018;159:21-26.
- Lewis SP, Seko Y, Joshi P. The impact of YouTube peer feedback on attitudes toward recovery from non-suicidal self-injury: An experimental pilot study. *Digit Health.* 2018 Jun 5;4:2055207618780499.
- Lewis SP, Seko Y. A Double-Edged Sword: A Review of Benefits and Risks of Online Nonsuicidal Self-Injury Activities. *J Clin Psychol.* 2016 Mar;72(3):249-62.
- Lewis SP, Baker TG. The possible risks of self-injury web sites: a content analysis. *Arch Suicide Res.* 2011;15(4):390-6.
- Lewis SP, Heath NL, St Denis JM, Noble R. The scope of nonsuicidal self-injury on YouTube. *Pediatrics.* 2011 Mar;127(3):e552-7.
- Luby J, Kertz S. Increasing Suicide Rates in Early Adolescent Girls in the United States and the Equalization of Sex Disparity in Suicide: The Need to Investigate the Role of Social Media. *JAMA Netw Open.* 2019;2(5):e193916
- Lupariello F, Curti SM, Coppo E, Racalbuto SS, Di Vella G. Self-harm Risk Among Adolescents and the Phenomenon of the "Blue Whale Challenge": Case Series and Review of the Literature. *J Forensic Sci.* 2019 Mar;64(2):638-642
- Marchant A, Hawton K, Burns L, Stewart A, John A. Impact of Web-Based Sharing and Viewing of Self-Harm-Related Videos and Photographs on Young People: Systematic Review. *J Med Internet Res.* 2021 Mar 19;23(3):e18048

- Marchant A, Hawton K, Stewart A, Montgomery P, Singaravelu V, Lloyd K, Purdy N, Daine K, John A. Correction: A systematic review of the relationship between internet use, self-harm and suicidal behaviour in young people: The good, the bad and the unknown. *PLoS One*. 2018 Mar 1;13(3):e0193937.
- McTernan N, Ryan F (2023). The Harmful Impact of Suicide and Self-Harm Content Online: A Review of the Literature. *National Suicide Research Foundation*
- Memon AM, Sharma SG, Mohite SS, Jain S. The role of online social networking on deliberate self-harm and suicidality in adolescents: A systematized review of literature. *Indian J Psychiatry*. 2018 Oct-Dec;60(4):384-392
- Miguel EM, Chou T, Golik A, Cornacchio D, Sanchez AL, DeSerisy M, Comer JS. Examining the scope and patterns of deliberate self-injurious cutting content in popular social media. *Depress Anxiety*. 2017 Sep;34(9):786-793.
- Minkkinen J, Oksanen A, Kaakinen M, Keipi T, Räsänen P. Victimization and Exposure to Pro-Self-Harm and Pro-Suicide Websites: A Cross-National Study. *Suicide Life Threat Behav*. 2017 Feb;47(1):14-26.
- Mitchell KJ, Wells M, Priebe G, Ybarra ML. Exposure to websites that encourage self-harm and suicide: prevalence rates and association with actual thoughts of self-harm and thoughts of suicide in the United States. *J Adolesc*. 2014 Dec;37(8):1335-44.
- Mok K, Jorm AF, Pirkis J. Suicide-related Internet use: A review. *Aust N Z J Psychiatry*. 2015 Aug;49(8):697-705.
- Mokkenstorm JK, Mérelle SYM, Smit JH, Beekman ATF, Kerkhof AJFM, Huisman A, Gilissen R. Exploration of Benefits and Potential Harmful Effects of an Online Forum for Visitors to the Suicide Prevention Platform in The Netherlands. *Crisis*. 2020 May;41(3):205-213.
- Mörch CM, Côté LP, Corthésy-Blondin L, Plourde-Léveillé L, Dargis L, Mishara BL. The Darknet and suicide. *J Affect Disord*. 2018 Dec 1;241:127-132.
- Moreno MA, Ton A, Selkie E, Evans Y. Secret Society 123: Understanding the Language of Self-Harm on Instagram. *J Adolesc Health*. 2016 Jan;58(1):78-84.
- Moss C, Wibberley C, Witham G. Assessing the impact of Instagram use and deliberate self-harm in adolescents: A scoping review. *Int J Ment Health Nurs*. 2023 Feb;32(1):14-29.
- Muslić L, Rukavina T, Markelić M, Musić Milanović S. Substance Use, Internet Risk Behavior, and Depressive Symptoms as Predictors of Self-harm Thoughts in Adolescents: Insights from the 2019 ESPAD Survey in Croatia. *Child Psychiatry Hum Dev*. 2023 Jul 25
- Nesi J, Burke TA, Bettis AH, Kudinova AY, Thompson EC, MacPherson HA, Fox KA, Lawrence HR, Thomas SA, Wolff JC, Altemus MK, Soriano S, Liu RT. Social media use and self-injurious thoughts and behaviors: A systematic review and meta-analysis. *Clin Psychol Rev*. 2021 Jul;87:102038.

- Orbyn A and Przybylski AK. The association between adolescent well-being and digital technology use. *Nat Hum Behav* 2019; 3:173-82
- Ortiz P, Khin Khin E. Traditional and new media's influence on suicidal behavior and contagion. *Behav Sci Law*. 2018 Mar;36(2):245-256.
- Ossa FC, Jantzer V, Neumayer F, Eppelmann L, Resch F, Kaess M. Cyberbullying and School Bullying Are Related to Additive Adverse Effects among Adolescents. *Psychopathology*. 2023;56(1-2):127-137
- Paul E, Mergl R, Hegerl U. Has information on suicide methods provided via the Internet negatively impacted suicide rates?. *PLoS One*. 2017;12(12):e0190136. Published 2017 Dec 28.
- Phillips JG, Diesfeld K, Mann L. Instances of online suicide, the law and potential solutions. *Psychiatr Psychol Law*. 2019;26(3):423-440.
- Pirkis J, Neal L, Dare A, Blood RW, Studdert D. Legal bans on pro-suicide web sites: an early retrospective from Australia. *Suicide Life Threat Behav*. 2009 Apr;39(2):190-3.
- Robertson L, Skegg K, Poore M, Williams S, Taylor B. An adolescent suicide cluster and the possible role of electronic communication technology. *Crisis*. 2012;33(4):239-45.
- Robinson J, Cox G, Bailey E, et al. Social media and suicide prevention: a systematic review. *Early Interv Psychiatry*. 2016;10(2):103-121
- Rodway, C., Tham, S., Richards, N., Ibrahim, S., Turnbull, P., Kapur, N., & Appleby, L.. Online harms? Suicide-related online experience: A UK-wide case series study of young people who die by suicide. *Psychological Medicine*, 2023, 53(10), 4434-4445.
- Ruder TD, Hatch GM, Ampanozi G, Thali MJ, Fischer N. Suicide announcement on Facebook. *Crisis*. 2011;32(5):280-2.
- Scherr S. Social media, self-harm, and suicide. *Curr Opin Psychol*. 2022 Aug;46:101311.
- Scheuerman MK, Jiang JA, Fiesler C, Brubaker JR. A Framework of Severity for Harmful Content Online. *Proc. ACM Hum.-Comput. Interact*. 5, CSCW2, Article 368
- Shafi RMA, Nakonezny PA, Romanowicz M, Nandakumar AL, Suarez L, Croarkin PE. Suicidality and self-injurious behavior among adolescent social media users at psychiatric hospitalization [published online ahead of print, 2020 Apr 27]. *CNS Spectr*. 2020;1-7.
- Sueki H. Does the volume of Internet searches using suicide-related search terms influence the suicide death rate: data from 2004 to 2009 in Japan. *Psychiatry Clin Neurosci*. 2011;65(4):392-394. doi:10.1111/j.1440-1819.2011.02216.x
- Sumner SA, Ferguson B, Bason B, Dink J, Yard E, Hertz M, Hilkert B, Holland K, Mercado-Crespo M, Tang S, Jones CM. Association of Online Risk Factors With Subsequent Youth Suicide-Related Behaviors in the US. *JAMA Netw Open*. 2021 Sep 1;4(9):e2125860

- Susi K, Glover-Ford F, Stewart A, Knowles Bevis R, Hawton K. Research Review: Viewing self-harm images on the internet and social media platforms: systematic review of the impact and associated psychological mechanisms. *J Child Psychol Psychiatry*. 2023 Aug;64(8):1115-1139
- The Lancet. Social media, screen time, and young people's mental health. *Lancet*. 2019;393(10172):611
- The Samaritans (2020). Managing self-harm and suicide content online: Guidelines for sites and platforms hosting user-generated content. Retrieved from: <https://www.samaritans.org/ireland/about-samaritans/research-policy/internet-suicide/guidelines-tech-industry/guidelines/> [Accessed November 3rd, 2020]
- The Samaritans. (2013). Media guidelines for reporting suicide. Retrieved from www.samaritans.ie/mediaguidelines [Accessed October 21st, 2020]
- Swedo EA, Beauregard JL, de Fijter S, Werhan L, Norris K, Montgomery MP, Rose EB, David-Ferdon C, Massetti GM, Hillis SD, Sumner SA. Associations Between Social Media and Suicidal Behaviors During a Youth Suicide Cluster in Ohio. *Journal of Adolescent Health*, 2020.
- Ueda M, Mori K, Matsubayashi T, Sawada Y. Tweeting celebrity suicides: Users' reaction to prominent suicide deaths on Twitter and subsequent increases in actual suicides. *Soc Sci Med*. 2017 Sep;189:158-166.
- Upadhyaya M, Kozman M. The Blue Whale Challenge, social media, self-harm, and suicide contagion. *Prim Care Companion CNS Disord*. 2022;24(5):22cr03314.
- Waszak PM, Górski P, Springer J, Kasprzycka-Waszak W, Duży M, Zagożdżon P. Internet searches for "suicide", its association with epidemiological data and insights for prevention programs. *Psychiatr Danub*. 2018;30(4):404-409.
- Winstone L, Mars B, Haworth CMA, Heron J, Kidger J. Adolescent social media user types and their mental health and well-being: Results from a longitudinal survey of 13-14-year-olds in the United Kingdom. *JCPP Adv*. 2022 Mar 10;2(2):e12071



VSPS Regulation
Coimisiún na Meán
2-5 Warrington Place
D02XP29
Ireland

By email VSPSregulation@cnam.ie

04/09/2023

Submission to the Call for Inputs: Online Safety Developing Ireland's First Binding Online Safety Code for Video-Sharing Platform Services

A Chara,

The HSE National Office for Suicide Prevention (NOSP) welcomes the opportunity to respond to the [Call for Inputs: Online Safety Developing Ireland's First Binding Online Safety Code for Video-Sharing Platform Services](#) (11th July 2023). Our background position is set out in this letter, followed by more specific answers to some of the questions provided in the Call for Inputs document.

Background

The HSE NOSP was established to strategically lead on suicide prevention efforts across the HSE and in collaboration with multiple partners. This work is underpinned by [Connecting for Life, Ireland's National Strategy to Reduce Suicide](#) (2015–2024). As a whole-of Government strategy, the HSE NOSP provides a strategic view of implementation progress within an implementation structure established in 2015. The Office fulfils a central role in this implementation structure and reports to the National Cross-sectoral Steering and Implementation Group (chaired by the Department of Health) on a quarterly basis.

The HSE NOSP also works directly with the non-governmental organisation sector – presently 21 agencies receive national funding from the Office to deliver on work aligned with the objectives and actions in Connecting for Life. Services and initiatives delivered across this diverse sector play a decisive role in advancing suicide and self-harm prevention, postvention and mental health promotion efforts in Ireland.

Connecting for Life sets out a vision of an Ireland where fewer lives are lost through suicide, and where communities and individuals are empowered to improve their mental health and wellbeing. The strategy has 69 actions, under 7 strategic goals.

- Goal 1: To improve the nation's understanding of and attitudes to suicidal behaviour, mental health and wellbeing
- Goal 2: To support local communities' capacity to prevent and respond to suicidal behaviour
- Goal 3: To target approaches to reduce suicidal behaviour and improve mental health among priority groups
- Goal 4: To enhance accessibility, consistency and care pathways of services for people vulnerable to suicidal behaviour
- Goal 5: To ensure safe and high-quality services for people vulnerable to suicide
- Goal 6: To reduce and restrict access to means of suicidal behaviour
- Goal 7: To improve surveillance, evaluation and high quality research relating to suicidal behaviour

Connecting for Life places a considerable emphasis on the need to 'engage and work collaboratively with the media in relation to media guidelines, tools and training programmes to improve the reporting of suicidal behaviour within broadcast, print and online media' (Objective 1.4). Four specific actions (1.4.1, 1.4.2, 1.4.3 and 1.4.4) detail a range of ways in which key stakeholders can encourage safer online environments, responsible media report and broadcasting of suicide-related content.

Suicide and self-harm

In preparation of this submission, the HSE NOSP has had deliberative discussions with relevant partners working in this area who have a specific interest in reducing the harmful impact of suicide and self-harm content online. These partners include the Department of Health (Mental Health Unit), Samaritans, Headline and the National Suicide Research Foundation (NSRF). As funder of Samaritans, Headline and the NSRF, the HSE NOSP has been supportive of their various initiatives in this broad area of work to date, and their separate present submissions to this Call for Inputs. The HSE NOSP is supportive of and endorses:

- Samaritans Ireland and [the Samaritans Media Guidelines for Ireland](#) – a range of guidance and information resources for media professionals, developed based on the evidence that certain types of media depictions, such as explicitly describing a method, sensational and excessive reporting, can lead to imitational suicidal behaviour among vulnerable people. Samaritans have also developed [Online Safety Guidelines](#), for sites and platforms hosting user-generated content.
- [Headline](#) (a project in Shine) – Ireland's national media programme for responsible reporting, and representation of mental ill health and suicide. Headline provides training, research, media

monitoring and support, for Irish media professionals across print, broadcast, and online platforms to reduce the effects of suicide contagion, and the stigma attached to mental ill health.

- [The National Suicide Research Foundation](#) (NSRF) – an independent, multi-disciplinary research unit that delivers research projects in suicide, self-harm and mental health. Support from the HSE NOSP ensures these projects can contribute to the surveillance, research, implementation, evaluation and the evidence base for strategic goals and actions of Connecting for Life. Of particular note, The Harmful Impact of Suicide and Self-harm Content Online: A Review of the Literature¹ sought to identify, review and summarise the literature and evidence on the impact of harmful suicide or self-harm content online, and to propose clearly defined descriptions of categories of online material that are considered to be harmful in relation to suicide and self-harm. This literature review has been revised and updated (2023).

Eating disorders

Connecting for Life, Ireland's National Strategy to Reduce Suicide, outlines priority groups for suicide prevention – groups for whom there is evidence of vulnerability to and increased risk of suicidal behaviour. The strategy also highlights risk factors of suicide that can be influenced by individual vulnerability or resilience, and these risk factors relate to the likelihood of a person developing suicidal behaviour. People with mental health problems, and notably people with eating disorders, have a heightened lifetime risk of, and vulnerability to, suicide.

In this context, the HSE NOSP would also take this opportunity to highlight the work of the [HSE National Clinical Programme for Eating Disorders \(NCP-ED\)](#), a collaborative initiative between the HSE, the College of Psychiatrists of Ireland, and [Bodywhys](#) (the Eating Disorders Association of Ireland), the national support group for people with eating disorders. Eating disorders have the highest mortality and morbidity of all of the mental health conditions², and it is estimated that they will affect between 1–4% of the population at some point in their lives. They are caused by a combination of genetic, biological and psychosocial factors and occur across gender, age, cultural, ethnic and socioeconomic groupings. Although not common, eating disorders result in very high psychosocial and economic cost to individuals, families, healthcare and society when not treated or treated ineffectively.³

Bodywhys asserts that while social media can be a great way to connect and provides opportunities to engage with areas of interest, it has also been highlighted as an additional pressure to body image. Research indicates that increased time spent online or on social media can impact negatively on body image. Social media posts tend to be about showing users best selves and the very best of their lives.

¹<https://www.hse.ie/eng/services/list/4/mental-health-services/connecting-for-life/publications/the-harmful-impact-of-online-content-a-literature-review.html>

² Arcelus, J., Mitchell, A. J., Wales, J., & Nielsen, S. (2011). Mortality rates in patients with anorexia nervosa and other eating disorders. A meta-analysis of 36 studies. *Archives of general psychiatry*, 68(7), 724–731. <https://doi.org/10.1001/archgenpsychiatry.2011.74>

³ <https://butterfly.org.au/>

Being bombarded with picture-perfect images of others can lead to a feeling of being 'not good enough'. Many people now also use filters and edit their photos and this can increase body image concerns as they might find it more difficult to accept their real-life selves. Editing of photos may also lead to an increased focus on the aspects of our appearance a user is not happy with, which may exacerbate body image concerns.

Bodywhys also developed [Guidelines for the Media](#) that outline broad principles (of avoiding specific details, avoiding sensationalising, covering 'celebrity' stories, the appropriate use of images, and on handling pro-anorexia websites) for media reporting, but could also be applied to general best practice in online or social content. The guidelines also contain information on best practice language and terminology related to eating disorders. Mindful use of language helps us to convey an understanding of the real needs of people affected by eating disorders and of the many challenges they face. Mindful use of language can also be a powerful tool in reducing stigmatisation thereby encouraging people towards seeking help.

For reference, in 2019, Bodywhys made a [submission on the 'Regulation of Harmful Content on Online Platforms and the Implementation of the Revised Audiovisual Media Services Directive'](#) that outlines their position on the risks and concerns related to eating disorders content online and on social media.

Future collaboration

Suicide prevention efforts require coordination and collaboration among multiple sectors of society, both public and private, including both health and non-health sectors such as education, labour, agriculture, business, justice, law, defence, politics and the media. These efforts must be comprehensive, integrated and synergistic, as no single approach can impact alone on an issue as complex as suicide.

The HSE NOSP looks forward to supporting the work of Coimisiún na Meán as their programme of work continues to develop in coming years and appreciates this present opportunity to impress the importance of reducing the harmful impact of suicide, self-harm and eating disorders content across a wide variety of platforms and online. We hope that consultative, collaborative and partnership approaches with stakeholders – particularly those working in health services and promotion – can continue.

Yours sincerely,

Mr John Meehan

HSE Assistant National Director, Mental Health Planning & Head of National Office for Suicide Prevention (NOSP), HSE Operations Planning



3. Online Harms

3.1 What online harms should the Code address?

Question 1: What do you think our main priorities and objectives should be in the first binding Online Safety Code for VSPS? What are the main online harms you would like to see it address and why?

The HSE National Office for Suicide Prevention (NOSP) is of the view that primary focus of an online safety code is to establish guidelines, rules, and practices that foster a safe, respectful, and inclusive online space for all users, this involves preventing and mitigating various forms of online harm.

Suicide and self-harm content

Online safety codes for online platforms typically have several main objectives, all aimed at ensuring a safe and secure online environment for users. These objectives may vary depending on the specific platform and the regulations in place but can include the following categories: Suicide and Self-harm Promotion, User Protection, Privacy, Content Moderation, Child Safety, Compliance with Laws and Regulations, Transparency and Accountability, Accessibility and Inclusivity, Addiction and Digital Well-being.

The HSE NOSP is of the view that the promotion of suicide and self-harm online is a key online harm that should be addressed in the forming of the online safety codes. A specific code that emphasises the harmful impact of pro suicide or self-harm material should be developed. This would assist in collective efforts to achieve objective 1.4 of Connecting for Life and other related efforts in this area.

Codes should encompass the following types of online content relating to suicide and self-harm:

- Information on how to hurt or kill oneself, including evaluations of different methods and rationale for each, and related questions and answers
- Chatrooms, forums or other material that encourages suicide or assists with suicide planning
- Suicide “pact” sites
- Images or videos that depict acts of suicide or self-harm, or locations/materials associated with such acts
- Material which promotes, facilitates or educates users on other suicidal behaviours e.g., behaviours that include planning for suicide, acquiring means to suicide, attempting suicide and suicide itself.

The HSE NOSP would also like to highlight the considerable attention that has been given to the possible role of social networks and the internet in contributing to self-harm and suicide contagion, predominantly in adolescence and youth.⁴ A study of associations between social media and suicidal

⁴ Becker K, Mayer M, Nagenborg M, El-Faddagh M, Schmidt MH. Parasuicide online: Can suicide websites trigger suicidal behaviour in predisposed adolescents? Nord J Psychiatry. 2004;58(2):111-4.

behaviours during a suicide cluster involving young people in the U.S. found that engagement with suicide cluster related social media was associated with increased suicide ideation and suicide attempts during a suicide cluster in Ohio.⁵

Internet sites and social media have been implicated in both inciting and facilitating suicidal behaviour. Private individuals can also readily broadcast uncensored suicidal acts and information that can be easily accessed through a wide variety of platforms.

It is also important to highlight that social media may also have a positive role in supporting people at risk for suicide.⁶ It is a belief of the HSE NOSP that any safety code produced should contribute to raising awareness about the potential dangers of suicide and self-harm promotion online by educating users about the consequences of sharing or consuming such content. Online platforms can collaborate with mental health experts and organisations to develop effective strategies for identifying and addressing harmful content. This may involve training content moderators to recognise warning signs and providing resources for users, and indeed moderators, who may need assistance.

While there is significant potential for harm from accessing pro-suicide material online (normalisation, triggering, competition, contagion) there is also the potential to exploit its benefits (crisis support, online help supports and information, reduction of social isolation, delivery of therapy, outreach) - and the design of safety codes should include appropriate consideration to achieving this balance. We should remember that all communities – including online communities – can play a critical role in suicide prevention. They can provide social support to vulnerable individuals and engage in follow-up care, fight stigma and support those bereaved by suicide.

Eating disorders

The HSE NOSP is also of the view that the promotion or encouragement of behaviour that characterises a feeding or eating disorder should be addressed in the forming of the online safety codes. With specific reference to eating disorders and according to Bodywhys⁷, the existence and activities of pro-anorexia websites and online content can be defined as those which tend to focus on the maintenance, promotion and encouragement of disordered eating behaviours and eating disorders. Typically, the websites operate without professional monitoring, supervision or formal guidance structures or support resources and channels. Terms used in this area include Pro-anorexia (pro-ana), Pro-bulimia (pro-mia) and Pro-eating disorder (pro-ED). Reasons for accessing these websites include:

- To pursue anorexia as a choice of 'lifestyle' through extreme thinness
- To manage issues that users feel are not adequately addressed in relationships outside of the internet

⁵ Swedo EA, Beauregard JL, de Fijter S, Werhan L, Norris K, Montgomery MP, Rose EB, David-Ferdon C, Masetti GM, Hillis SD, Sumner SA. Associations Between Social Media and Suicidal Behaviors During a Youth Suicide Cluster in Ohio. *Journal of Adolescent Health*, 2020.

⁶ Fu KW, Cheng Q, Wong PW, Yip PS. Responses to a self-presented suicide attempt in social media: a social network analysis. *Crisis*. 2013 Jan 1;34(6):406-12.

⁷ <https://www.bodywhys.ie/wp-content/uploads/2019/07/Bodywhys-online-safety-FINAL.pdf>

- To seek support from others with similar beliefs and experiences
- To seek reinforcement and a sense of community
- To seek support due to a lack of understanding and feeling marginalised from traditional support structures
- To exchange messages as a form of emotional support
- To cope with stigma and write online postings as a form of self-expression
- To maintain a concealed identity, including from family and friends.

Concerning behaviours discussed on these sites may include:

- How to maintain or initiate eating disorder behaviours and how to resist treatment or recovery
- How to obtain and use weight loss medications
- How to conceal anorexia from family members
- How to behave in social situations involving food, particularly when interacting with people who do not have an eating disorder
- Information on weight loss strategies, commonly known as tips and tricks
- Diet challenges and competitions
- Praise for the denial of nourishment
- Disguising evidence of and how to induce vomiting, the sharing of personal photographs of emaciation in order to seek approval and validation from peers.

3. Online Harms

3.1 What online harms should the Code address?

Question 2: What types of online harms do you think should attract the most stringent risk mitigation measures by VSPS? How could we evaluate the impact of different types of harms e.g. severity, speed at which harm may be caused? Is there a way of classifying harmful content that you consider it would be useful for us to use?

What types of online harms do you think should attract the most stringent risk mitigation measures by VSPS?

It is the view of the HSE NOSP that harmful content relating to suicide and self-harm should attract stringent risk mitigation measures by VSPS, in addition to content that promotes or encourages behaviour that characterises a feeding or eating disorder.

As outlined in the response to question 1 the HSE NOSP is of the view that the promotion, facilitation or education of suicide and self-harm methods online is a key online harm that should be directly addressed. Therefore a specific code addressing pro suicide or self-harm content should be developed and assigned stringent risk mitigation measures. This would assist in collective efforts to achieve objective 1.4 of Connecting for Life and other related efforts in this area.

The HSE NOSP draws attention to increases that have been noted in changes in the accessibility of explanatory information on methods of suicide. For example, in a study analysing changes between 2007 and 2017, over 54% of hits from search terms related to suicide, contained information about new high-lethality methods of suicide.⁸

Additional research from the University of Bristol in 2015 found that in a population survey of 21 year olds, of the 248 participants who had made suicide attempts (6% of the overall sample), almost three quarters reported some kind of suicide-related internet use at some point in their lives.⁹ One in five had accessed sites giving information on how to hurt or kill oneself, though most of these had also visited help-sites. In a clinical sample of over 1,500 patients who presented to hospital following a suicide attempt, 8% said they had used the internet in connection with their attempt. This percentage was higher for younger patients (12% of those aged 16-24 years) and those who had self-harmed with high suicidal intent (24%). For most of those interviewed in the clinical sample, the main purpose for going online was to research methods of suicide, sometimes in great depth. While researching methods of suicide online, did not always lead to action, it made individuals vulnerable by validating their feelings, legitimising suicide as a course of action, and providing knowledge about methods of suicide. Half of

⁸Gunnell D, Derges J, Chang S-S, Biddle L. Searching for suicide methods. *Crisis*, 2015. 36(5): 325-331. <http://dx.doi.org/10.1027/0227-5910/a000326>

⁹Mars B, Heron J, Biddle L, Donovan J, Holley R, Piper M, Potokar J, Wyllie C, Gunnell D. Exposure to, and searching for, information about suicide and self-harm on the Internet: Prevalence and predictors in a population based cohort of young adults. *Journal of Affective Disorders*, 2015. 185 p 239-245. <http://dx.doi.org/10.1016/j.jad.2015.06.001>

those interviewed in the clinical sample planned and carried out a suicide method, based on their online research; some had purchased materials online. However, in some instances, information about methods discovered online was found to be ‘off putting’, causing some individuals to rule out particular methods of suicide.

A more recent (2017) review of the evidence of the relationship between internet use, self-harm and suicidal behaviour in young people, highlights that this relationship is particularly associated with internet addiction, high levels of internet use, and websites with self-harm or suicide content.¹⁰

How could we evaluate the impact of different types of harms e.g. severity, speed at which harm may be caused?

Evaluating the impact of different types of online harms involves assessing various factors, and the HSE NOSP would like to specify the following markers in relation to the curation of harmful content relating to suicide and self-harm, and eating disorders. In this context, the risk of potential harm to – or suicide among – vulnerable users and communities, can be high.

It is important to consider the severity of the harm; e.g., whether it could lead to physical harm or danger to individuals, while also considering the psychological impact, assessing the potential psychological impact on individuals, such as emotional distress, anxiety, depression, or trauma. In the context of suicide or self-harm it is important to always be mindful of the vulnerabilities of the users, particularly those who may be experiencing suicidal ideation or perhaps bereaved by suicide.

The design of codes should make consideration of how quickly harmful content can spread through social media, messaging platforms, or other online channels, potentially reaching a large audience in a short period. This is of particular concern in reference to self-harm and suicide contagion as outlined in question 1 of this document.

There should also be scope to review the real-world impact of such harms; i.e., assessing whether the harm can lead to tangible real-world consequences, such as physical harm, instances of self-harm or a death by suicide. It is also important to consider whether the harm poses a threat to public safety. Further to this, attention should be given to whether the harmful content has the potential to go viral or be widely shared, thus amplifying its impact and reach.

The design of codes should also make consideration of the feasibility and effectiveness of implementing measures to mitigate the harm, such as content moderation, reporting mechanisms, or algorithmic adjustments. It is also important to determine if the harm is more likely to affect children, minors, or other vulnerable user groups who may be less equipped to handle or discern harmful content.

Is there a way of classifying harmful content that you consider it would be useful for us to use?

¹⁰ Marchant A, Hawton K, Stewart A, Montgomery P, Singaravelu V, Lloyd K, et al. (2017) A systematic review of the relationship between internet use, self-harm and suicidal behaviour in young people: The good, the bad and the unknown. PLoS ONE 12(8): e0181722. <https://doi.org/10.1371/journal.pone.0181722>

Classifying harmful content can be a useful approach to better understand, categorise, and address the various types of online harms. This classification can help platforms, policymakers, and researchers develop targeted strategies for prevention, moderation, and user education. The HSE NOSP would be supportive of work to develop a classification system for harmful online content related to suicide and self-harm, and eating disorders.

The following list, taken from *The Harmful Impact of Suicide and Self-Harm Content Online: A Review of the Literature* aims to clearly define descriptions of categories of online material that are considered to be harmful in relation to suicide and self-harm. In line with these aims, this answer is segmented into ten main sections, categorised by the following types of online content:

1. Online information sources (websites used to inform method)
2. Search engines
3. Social networks
 - a. Facilitate access to potentially harmful information
 - b. Facilitate contagion
 - c. Normalising self-harm and suicide
 - d. Increased risk following celebrity suicide
 - e. Facilitate cyberbullying
 - f. Suicide notes
4. Online imagery and videos
5. Online forums/message boards
6. Pro-suicide and self-harm sites
7. Online suicide 'games'
8. The 'Darknet'
9. Livestream suicide / cybersuicide
10. Online suicide 'pacts'¹¹

¹¹ Niall McTernan and Fenella Ryan, *The Harmful Impact Of Suicide And Self-Harm Content Online: A Review Of The Literature*. National Suicide Research Foundation, Ireland

3. Online Harms

3.1 What online harms should the Code address?

Question 3: Do you have reports, academic studies or other relevant independent research that would support your views? If you do, please share them with us with links to relevant reports, studies or research.

Suicide, self-harm

[The Harmful Impact of Suicide and Self-harm Content Online: A Review of the Literature \(NSRF, 2019 and 2023\)](#). This report contains an extensive bibliography and references list.

General Mental Health, including Eating Disorders

[Can the Metaverse Be Good for Youth Mental Health? Youth-Centered Strategies...](#) (jedfoundation.org)

[Online advertising and eating disorders - Beat](#) (beateatingdisorders.org.uk)

[New research from Butterfly Foundation highlights impact of social media - Butterfly Foundation](#)

[The impact of digital experiences on adolescents with mental health vulnerabilities | Media@LSE](#)

[Investigating-Risks-and-Opportunities-for-Children-in-a-Digital-World.pdf](#) (unicef-irc.org)

[CCDH-Deadly-by-Design_120922.pdf](#) (counterhate.com)

[Social Media and Youth Mental Health](#) (hhs.gov)

[American Psychological Association Health Advisory on Social Media Use in Adolescence](#) (apa.org)

[Social Media And Apps-FREED.pdf](#) (freedfromed.co.uk)

4. Overall Approach to the Code

4.1 How prescriptive or flexible should the Code be?

Question 4: What approach do you think we should take to the level of detail in the Code? What role could non-binding guidance play in supplementing the Code?

It is the opinion of the HSE NOSP that 'Option 3 – A mixed approach' is the best way to approach the level of detail required in the code.

Non-binding guidance can play a valuable role in supplementing an online safety code by providing additional context, practical advice, and best practices for both platform operators and users. While an online safety code sets the foundation for expected behaviours and standards, non-binding guidance can offer more detailed insights into how to effectively implement and adhere to those standards.

Non-binding guidance can elaborate on the principles and rules outlined in the safety code, helping users and content creators better understand the intended behaviours and standards. This clarity can reduce ambiguity and prevent unintentional violations.

Guidance documents can also offer step-by-step instructions and practical tips on how to create and share content in a safe and responsible manner. This can allow users to make informed decisions and contribute positively to the online community.

Guidance can provide strategies for identifying and responding to harmful content, as well as tips for protecting oneself and others from potential risks. It can also provide content moderators with more context on how to interpret and apply the safety code ensuring consistent and fair enforcement of the rules.

In the context of mental health promotion and suicide prevention the HSE NOSP is of the opinion that the accompaniment of all codes with appropriate guidance, campaigns and educational initiatives will be helpful in ensuring a consistent and collaborative approach to fostering change.

This will also assist stakeholders and agencies to participate, endorse and promote the work of Coimisiún na Meán, and better understand the application of codes in the 'real-world' and their impact on the people who they work to support.

5. Measures to be taken by Video-Sharing Platforms

5.1 Online Safety Features for Users

Question 7: To what extent, if at all, should the Code require VSPS providers to take measures to address content connected to video content?

Question 9: How should we ask VSPS providers to introduce and design a flagging mechanism in the Code? How can we ensure that VSPS providers introduce the mechanism in a user-friendly and transparent way? How should we ask VSPS Providers to report the decisions they've made on content after it has been flagged? To what extent should we align the Code with similar provisions on flagging in the DSA?

Question 10: What requirements should the Code include about age verification and age assurance? What sort of content should be shown by default to users who are logged out or in private browsing mode and whose age cannot be verified or assured? What evidence is there about the effectiveness of age estimation techniques? What current practices do you regard as best practice? Where accounts are not age verified should default privacy settings be used, should content default to universal content and should contact by others be more limited?

The HSE NOSP is broadly supportive of the ten measures to protect the general public and children from online harms that are already set out in Article 28b.3 of the AVMSD. In the context of harmful suicide, self-harm and eating disorders content online, the establishment of transparent and user-friendly mechanisms for users to report or flag content, and for VSPS providers to explain to users the same, is particularly important.

In addition:

- Consideration should be given to the evidence of the effectiveness¹² and dependability¹³ of generalised 'trigger warnings'.
- Comprehensive information on help, supports and services should accompany flagging mechanisms for users. This information should be:
 - Appropriately aligned with the nature and severity of the content, and sophisticated enough to return local information, or time-specific information. For example, in critical or emergent incidents, signposting to emergency, out-of-hours local services.
 - Routinely reviewed and validated with relevant support services and accurate at all times.

¹² Bridgland, V.M., Jones, P.J. and Bellet, B.W., 2022. A meta-analysis of the efficacy of trigger warnings, content warnings, and content notes. *Clinical Psychological Science*, p.21677026231186625.

¹³ Moreno MA, Ton A, Selkie E, Evans Y. Secret Society 123: Understanding the Language of Self-Harm on Instagram. *J Adolesc Health*. 2016 Jan;58(1):78-84.

- More integrated real-time connections or solutions could also be designed and established between VSPS providers and appropriate 24-hour support service providers. For example, the establishment of integrated access to text, helpline or emergency services.

Suicide and self-harm content online (that is harmful or otherwise) can arise and propagate quickly, therefore emphasis should be given to ensure such mechanisms are real-time, efficient and responsive, in particular when incidents have occurred locally, nationally or internationally. In these instances, the potential for severe, rapid and real-world harm is considerable. For example, when a public figure or high-profile personality has died by (suspected) suicide, or when a community has experienced a loss or multiple losses.

The HSE NOSP recommends that appropriate working partnerships are formed between relevant agencies (for example, in health services) and VSPS providers, to inform how they design, prioritise and address content moderation issues and potential timescales for moderation decisions and action. These working partnerships could be grouped or assigned to themes, domains or categories of harmful online content as established.

The establishment of codes and their application may also present opportunities for more sophisticated integrated responses to death(s) by suicide, from health services and communities. For example, [Developing a Community Response to Suicide](#) (a resource to guide those developing and implementing an Inter-Agency Community Response Plan for incidents of suspected suicide, particularly where there is a risk of clusters and/or contagion) outlines how a wide variety of agencies should work together to respond to suicide, and potentially provides forums locally and nationally, for VSPS providers to support and participate in these preventative efforts.

The establishment of a consistent mechanism or requirement for VSPS providers to report routinely on their content moderation metrics or decisions, would be particularly beneficial. This would help to enhance a broader understanding – across all sectors – of the issues arising, and assist research and building the evidence base for how the harmful impact of suicide and self-harm content online can be minimised. It will assist suicide and self-harm service providers and policy makers alike, to design and frame their own objectives and actions in this area of work.

5. Measures to be taken by Video-Sharing Platforms

5.2 Terms and Conditions, Content Moderation and Complaints

Question 14: How should we ask VSPS providers to address online harms in their terms and conditions in the Code, including the harms addressed under Article 28b? How should key aspects of terms and conditions be brought to users' attention? What examples are there of best practice in relation to terms and conditions including content moderation policies and guidelines?

Online harms as described in terms and conditions, and in moderation policies and guidelines, should be clearly presented by VSPS providers, in readily available accessible formats. Information on harmful content related to suicide, self-harm and eating disorders, should be accompanied by user friendly information on why the content might be harmful, and aim to improve users own understanding and encourage personal responsibility in this area.

Explicit information on what content is prohibited, and on sanctions that users need to be aware of, should be readily available from VSPS providers. As a priority, categorisation and prohibition of content that has potential to cause severe physical harm or psychological impact, should be clearly defined and accessible to users.

In the context of suicide and self-harm, this might include content that:

- Provides information on how to hurt or kill oneself, including evaluations of different methods and rationale for each, and related questions and answers
- Promotes chatrooms, forums or other material that encourages suicide or assists with suicide planning
- Promotes suicide “pact” sites
- Includes livestreams or a person attempting suicide
- Promotes other suicidal behaviours e.g., behaviours that include planning for suicide, acquiring means to suicide, attempting suicide and suicide itself.

In the context of eating disorders, this might include content that:

- Provides information on how to maintain or initiate eating disorder behaviours and how to resist treatment or recovery
- Provides information on how to obtain and use weight loss medications
- Provides information on how to conceal anorexia from family members
- Provides information on how to behave in social situations involving food, particularly when interacting with people who do not have an eating disorder
- Provides information on how weight loss strategies, commonly known as tips and tricks
- Encourages diet challenges and competitions
- Provides praise for the denial of nourishment

- Promotes the disguising evidence of and how to induce vomiting, the sharing of personal photographs of emaciation in order to seek approval and validation from peers.

Question 15: How should we ask VSPS providers to address content moderation in the Code? Are there any current practices which you consider to be best practice? How should we address automated content detection and moderation in the Code?

Safe and effective content moderation (automated or otherwise) of suicide, self-harm and eating disorders content online is of utmost importance. Given the potential for real-time harm, particularly to vulnerable groups, targeted and proportionate obligations should exist for VSPS providers to monitor such content.

It is essential that content moderators – who are routinely exposed to potentially harmful content online – should be appropriately trained, supported and supervised in their work, to ensure their own safety and wellbeing in the context of such emotive, sensitive and sometimes distressing and traumatic content.

Favourable consideration should also be given to mandating VSPS providers to prioritise or escalate requests, from nominated or assigned – where applicable – subject matter experts, regulators, public bodies, or health services.

For example, the HSE NOSP would be supportive of engaging directly with VSPS providers, to establish protocols or procedures that would assist real-time health or community detection of and responses to critical incidents or cases of (suspected) suicide, at local levels. [Developing a Community Response to Suicide](#) (a resource to guide those developing and implementing an Inter-Agency Community Response Plan for incidents of suspected suicide, particularly where there is a risk of clusters and/or contagion) outlines how a wide variety of agencies should work together to respond to suicide, and potentially provides forums locally and nationally, for VSPS providers to support and participate in these preventative efforts.

5. Measures to be taken by Video-Sharing Platforms

5.3 Possible Additional Measures and Other Matters

Question 19: How do you think that cooperation with other regulators and bodies can help us to implement the Code for VSPS?

In the context of suicide and self-harm prevention, and supporting people with eating disorders, a wide range of agencies, communities and statutory/non-statutory bodies is required to work effectively, in partnership and with a shared understanding of evidence-based prevention intervention and postvention responses. VSPS providers should be required to cooperate with health agencies and providers (where relevant) to ensure that codes can be understood and implemented effectively. Significant opportunities will exist for aligning for example, mental health/suicide prevention supports, public health information and campaigns, with the codes and mechanisms arising across different VSPS providers.

Question 20: What approach do you think we should take in the Code to address feeds which cause harm because of the aggregate impact of the content they provide access to? Are there current practices which you consider to be best practice in this regard?

Many vulnerable users who access harmful suicide, self-harm or eating disorder content online, may do so at a particular time of crisis or vulnerability. For most, these times pass, and with the right support, service or intervention, move to a place of wellness in time. Therefore, it is essential that consideration is given to the potential for harm, of aggregate suicide or self-harm content over time that has been recommended to a user.

For example, the internet is frequently used to obtain information about methods of suicide and self-harm¹⁴. Several respondents to a UK hospital-based qualitative study admitted to intentionally seeking information about methods when planning their attempt — predominantly from the internet¹⁵. As a priority, algorithms should therefore be designed to minimise or eliminate the recurrence and further recommendations of such harmful content, which can have potentially severe consequences. Instead, algorithms should be designed under principles of harm reduction and recovery and should promote trusted and validated support content to vulnerable users, as appropriate to the severity of the content that a user originally accessed.

¹⁴ Robert A, Suelves JM, Armayones M, Ashley S. Internet use and suicidal behaviors: internet as a threat or opportunity? *Telemed J E Health*. 2015 Apr;21(4):306-11

¹⁵ Biddle L, Gunnell D, Owen-Smith A, Potokar J, Longson D, Hawton K, Kapur N, Donovan J. Information sources used by the suicidal to inform choice of method. *J Affect Disord*. 2012 Feb;136(3):702-9.



MPIL RESPONSE TO CALL FOR INPUTS: ONLINE SAFETY; DEVELOPING IRELAND'S FIRST ONLINE SAFETY CODE FOR VIDEO SHARING PLATFORM SERVICES

4 September, 2023

FAO: Laura Forsythe (By email: vspsregulation@cnam.ie)

Dear Laura

Meta Platforms Ireland Limited (**MPIL**) welcomes the opportunity to make submissions in response to Coimisiún na Meán's (**an Coimisiún**) call for inputs on the development of Ireland's first Online Safety Code (the **Code**) (the **CFI**).

We note that an Coimisiún designated video-sharing platform services (**VSPS**) as a category of services under the Online Safety and Media Regulation Act 2022 (**OSMR**) on 14 August 2023, and that the Code will focus on this category of services.

MPIL is committed to protecting our users' voices and helping them connect and share safely. We want Facebook (**FB**) and Instagram (**IG**) to be safe and enjoyable places for our users to engage and connect with people and interests that are important to them. In order to achieve this, Meta has invested significant resources - both human and technology - to ensure that our platforms are as safe as possible.

We have been calling for the implementation of the revised Audiovisual Media Services Directive since it became EU law in 2018 and we welcome the fact that progress is now being made in that regard. We believe that the Directive's implementation will contribute to the development of a harmonised approach to harmful online content in the European Union, complementing the Digital Services Act (**DSA**) and other existing and planned Union law. A common EU approach is in the interest of all stakeholders, including users.

We welcome an Coimisiún's intention to design the Code to minimise the potential for conflict and maximise the potential for synergies with the DSA. We strongly agree with this sentiment and the need to ensure that the Code does not conflict with the DSA (nor any other EU regulatory regimes).

To this end, when drafting the Code, an Coimisiún should keep in mind the importance of the uniform application of the DSA's harmonised rules to "*put an end to fragmentation of the internal market*" and "*ensure legal certainty*" (see Recital 4 DSA) and that Member States not adopt national measures dealing with requirements addressing the dissemination of illegal content online, as this is expressly recognised as an area which should be "*fully*" harmonised under the DSA (see Recital 9 DSA).


Confidentiality

We look forward to discussing confidentiality protocols with an Coimisiún in due course to ensure that this data is afforded appropriate protection. In the interim, we respectfully request that the contents of this submission not be disclosed outside an Coimisiún without our prior engagement.

We thank an Coimisiún for the opportunity to provide comments on the Call for Inputs, and we hope that our comments will assist an Coimisiún in carrying out its regulatory functions. We are available to discuss any aspect of the below with an Coimisiún at any stage, and we look forward to regular engagement with an Coimisiún as it develops its first Online Safety Code.

Yours sincerely

Meta Platforms Ireland Limited

APPENDIX

SECTION 3 - ONLINE HARMS: WHAT ONLINE HARMS SHOULD THE CODE ADDRESS?

Question 1: What do you think our main priorities and objectives should be in the first binding Online Safety Code for VSPS? What are the main online harms you would like to see it address and why?

What do you think our main priorities and objectives should be in the first binding Online Safety Code for VSPS?

An Coimisiún's main priorities and objectives for the Code should be clarity, proportionality and the avoidance of duplicative regulatory requirements in the regulation of VSPS.

As an Coimisiún is aware, VSPS will be required to implement online safety measures under the EU Digital Services Act (the **DSA**) and some VSPS will be subject to the additional requirements applicable to very large online platforms (**VLOPs**), including Facebook and Instagram (please see response to **Question 6** for further information). In developing the Code, an Coimisiún should therefore prioritise **consistency with existing and future regulatory requirements** applicable to VSPS. Additionally, we would note that many VSPS, including those provided by MPIL, already have policies and practices in place to tackle harmful and illegal online content and are actively implementing online safety measures to ensure that users have as safe and as enjoyable an experience as possible on their services. Accordingly, the Code should also recognise, and take account of, the existing efforts of VSPS in relation to online safety, as these will undoubtedly assist VSPS in achieving the objectives of the Code.

We consider that an Coimisiún's objectives for the Code should include establishing a "baseline" of measures which VSPS providers commonly have in place and which form part of industry best practice and existing regulatory requirements. This will ensure that the Code is consistent with and accounts for existing practice and relevant regulatory regimes. An Coimisiún could then build on these "baseline" measures where it considers that existing requirements are not sufficient to meet the objectives sought to be achieved by the Code. This would also align with the regime envisaged under the Audiovisual Media Services Directive (**AVMSD**), which we understand was intended to be an iterative and evolving regulatory framework, rather than a regime which could be captured in a single code. This would also be in line with other industry codes, such as the the New Zealand Code of Practice for Online Safety and Harms (the **CPOSH**)¹

Additionally, given that not all VSPS are the same and an Coimisiún's intention is to adopt one Code (at least initially) that will apply to all VSPS, flexibility will be crucial to ensuring that the measures VSPS are required to implement can be applied effectively for each service. We believe that the most appropriate way to achieve flexibility under the Code is to adopt a **principles-based approach**. The

¹The CPOSH aims to provide best practices for a broad range of products and services, serving diverse and different user communities with different use cases and concerns. As such, it provides flexibility for potential Signatories to innovate and respond to online safety and harmful content concerns in a way that best matches their risk profiles, as well as recalibrate and shift tactics in order to iterate, improve and address evolving threats online in real-time (see p. 1 here: <https://netsafe.org.nz/wp-content/uploads/2021/12/Aotearoa-New-Zealand-Code-of-Practice-for-Online-Safety-and-Harms-public-feedback-draft.pdf>)

Code should operate with guiding principles, as is customary under other codes of practice, like the CPOSH². This will allow for flexibility in compliance solutions and the ability to iterate compliance measures as new developments occur and in light of relevant factors, e.g. nature of service, user base, existing measures in place, etc. Such an approach aligns with the principles-based approach to harmful online content taken in the AVMSD and is critical to ensuring that measures imposed are “practicable and proportionate” as required by the AVMSD (see, for instance, Article 28b(3), which recognises that a range of complex factors need to be taken into account in determining whether measures are appropriate³).

This approach is also consistent with previous statements by the Broadcasting Authority of Ireland (BAI) and in the General Scheme of the OSMR. For example, in its Submission to the Department of Communications, Climate Action & Environment Public Consultation on the Regulation of Harmful Content on Online Platforms and the Implementation of the Revised Audiovisual Media Service Directive (**BAI Submission**), the BAI noted that the revised AVMSD advocated for “a principles-based approach to protection” whereby “high level rules and principles” would be drawn up and VSPS would be “obliged to follow a principles-based common code”.

As noted in the General Scheme: “in overall terms, it’s important to note that the Media Commission would develop, in the first instance, high level principle based codes governing standards and practices. Designated online services are then required to develop measures to meet the principles set out in the high level codes that apply to them..... This approach provides for the Media Commission, through learned experience, to develop more detailed and tailored codes in certain discrete areas as standardised best practices emerge. It also provides for a quasi-continuous process of improving measures taken by online services to meet the requirements of the high-level codes through ongoing engagement and assessment by the Media Commission”.

The intention was clearly that the regulation of VSPS would allow for an iterative and evolving regulatory framework, taking a principles-based approach rather than a prescriptive approach seeking to apply rigid criteria to dynamic platforms in a one-size-fits-all manner. This recognises that not all service providers are the same - flexibility is critical in terms of optionality in mitigations and tools. Indeed in the UK, with 2+ years’ experience, Ofcom recognises that the risks posed by content “is highly contextual and dependent on a range of factors, including the age and demographic of users” (paragraph 2.8 of Ofcom’s 2023 [User Policies Report](#)). A principles-based approach, allowing the platform to ultimately choose the most effective measures, is also consistent with the approach under the DSA and the EU’s Terrorist Content Online Regulation.

To ensure the Code’s priorities and objectives are workable and effective, they should be

² The CPOSH commits signatories to a set of guiding principles, commitments, outcomes and measures that are focused on seven safety and harmful content themes - 1) child sexual exploitation and abuse; 2) bullying or harassment; 3) hate speech; 4) incitement of violence; 5) violent or graphic content; 6) misinformation; and 7) disinformation - which Netsafe and the Signatories believe are of great concern for Aotearoa New Zealand internet users. This makes the Code much broader than other existing industry codes, and commits signatories to provide transparency about their policies, processes and systems (see p. 2 here: <https://netsafe.org.nz/wp-content/uploads/2021/12/Aotearoa-New-Zealand-Code-of-Practice-for-Online-Safety-and-Harms-public-feedback-draft.pdf>).

³ Article 28b(3) AVMSD recognises that a range of complex factors need to be taken into account in determining whether measures are appropriate, including: a) the size and nature of the video-sharing platform service; b) the nature of the material in question; c) the harm the material in question may cause; d) the characteristics of the category of persons to be protected (for example, under-18s); e) the rights and legitimate interests at stake, including those of the person providing the video-sharing platform service and the persons having created or uploaded the material, as well as the general public interest.

evidence-based and rooted in research. In this regard, we would encourage an Coimisiún to draw upon the large body of research conducted by other regulators globally, for example, Netsafe in New Zealand and OFCOM in the UK (see **Question 3** below). We are aware that a principles-based approach to regulation has been highly effective in other jurisdictions and are of the opinion that aligning the regulation of VSPS in Ireland with other jurisdictions would result in a more effective regulatory regime overall, as online harms are, by their very nature, global, and so regulation of online harms should be reflective of this.

What are the main online harms you would like to see it address and why?

The CFI clarifies that an Coimisiún intends for the Code to complete the transposition of Article 28b of the AVMSD in Ireland. We agree that this should be an Coimisiún's focus under the Code, as transposition of the AVMSD is currently an urgent priority for Ireland.

The AVMSD identifies the categories of harm in respect of which VSPS should take appropriate measures and, as such, we consider the main online harms that should be addressed in the Code are the categories set out at Article 28b (1)(a), (b) and (c) of the AVMSD, i.e., protection of minors, incitement to violence or hatred and specific illegal harms (terrorism, CSAM, offences concerning racism and xenophobia). This will ensure clarity under the Code and in respect of an Coimisiún's regulatory expectations.

For the avoidance of doubt, the VSPS Code should only apply to those categories of content specified in Article 28b(1) of the AVMSD. Article 28b(6) permits Member States to impose 'measures' that are more detailed or stricter than those referred to in paragraph 3 but does not permit Member States to include additional categories of content. Accordingly, measures applying to additional categories of content would not benefit from the country-of-origin principle under the AVMSD and so would create legal uncertainty for an Coimisiún, regulated entities and users.

If the Code were to go beyond the requirements of the AVMSD, we are of the view that this would lead to an increased risk that the Code may conflict with other online safety legislation, most notably the DSA, and would result in an uncertain and duplicative regulatory regime for VSPS.

Question 2: What types of online harms do you think should attract the most stringent risk mitigation measures by VSPS? How could we evaluate the impact of different types of harms e.g. severity, speed at which harm may be caused? Is there a way of classifying harmful content that you consider it would be useful for us to use?

Protecting users from different types of harmful and illegal content is something that we take very seriously. We want all users to have a safe and enjoyable experience when they use our products. We have developed a range of measures and solutions to tackle these different types of harm. Our experience is that there are a number of consistent measures that are appropriate and useful across the board i.e. those measures are suitable regardless of the type of harm. Whereas, some additional measures which are more tailored, may be appropriate to take for specific harm types (described further below).

It is important, in the development of the Code, to recognise how VSPS commonly address and mitigate issues surrounding harmful and illegal content based on their experience including, the nature of the services.

At Meta, first and foremost, we maintain globally applicable standards – Facebook’s Community Standards⁴ and Instagram’s Community Guidelines⁵ – that define what is and isn’t allowed on our services. These standards apply uniformly to content worldwide and are integral to protecting expression and enhancing personal safety on our services. Many of our standards focus on content that is or is likely to be illegal though they do not map to specific laws, which vary significantly around the world. Our standards address the types of potentially harmful content that are of greatest concern or are seen most commonly on our platforms. In addition, our standards prohibit a wide range of objectionable or harmful content that is not necessarily illegal, including content that is considered graphic violence, spam, misinformation or bullying.

Our standards are created by global teams with a wide array of backgrounds and expertise, including those who have dedicated their careers to issues like child safety, hate speech, and terrorism. We regularly seek input from outside experts and organisations to help balance the different perspectives that exist on free expression and safety, and to better understand the potential impacts of our policies on different communities globally. Our reviewers enforce these standards using comprehensive guidelines, in an effort to ensure that decisions are as consistent as possible.

Second, across our platforms we action⁶ millions of pieces of content per day, through both human review and automation. In most cases, this happens automatically, with technology such as artificial intelligence working behind the scenes to detect and remove violating content⁷. To track our progress and demonstrate our continued commitment to making Facebook and Instagram safe and inclusive we publish the Community Standards Enforcement Report (**CSER**) on a quarterly basis⁸.

We also take a number of steps to assess and mitigate the risk of harm to our users from content that violates our standards (which, as noted above, overlaps with various types of illegal content) through the entire product development cycle. Before launch, new products generally go through an Integrity Review, a cross-functional (XFN) process which evaluates product changes on integrity criteria prior to launch. In this process, products are reviewed against a set of integrity standards to help us provide a positive experience for users. As part of this process, we systematically and repeatedly bring together experts from across the company, including data scientists, safety experts and engineers. The process helps us identify and anticipate potential abuses and build in mitigations by design, prior to a product being launched. After launch, we continue to monitor the potential impact of our products.

Another way we try to help our users to stay safe, including in relation to illegal content, is by


⁴ See <https://transparency.fb.com/en-gb/policies/community-standards/>

⁵ See <https://help.instagram.com/477434105621119>

⁶ Taking action on content could include removing a piece of content from Facebook or Instagram, covering photos or videos that may be disturbing to some audiences with a warning, or disabling accounts.

⁷ See <https://transparency.fb.com/en-gb/enforcement/>

⁸ See <https://transparency.fb.com/reports/community-standards-enforcement/>



providing a suite of in-app safety tools and privacy and security features available to all users service-wide. By way of example, (1) we have put in place safeguards against child exploitation, such as photo-matching technologies that help detect, remove, and report the sharing of images and videos that exploit children; (2) likewise, we deploy technology to detect and remove non-consensual intimate imagery, using image processing and media match software; (3) as part of our efforts to keep people safe and address content that may be detrimental to the wellbeing of teens, we use technical measures to proactively find and remove harmful suicide and self-harm content, which enables us to look for posts that likely break Facebook and Instagram’s rules around suicide and self-harm and make them less visible by down ranking them, and where we are confident that the content breaks the rules, remove that content; (4) we deploy numerous safeguards for teen interactions on Instagram’s direct messaging features, such as restricting adults from direct messaging teens who don’t follow them, as well as sending safety notice prompts to encourage teens to be cautious in conversations, a teen and adult are already connected/following; (5) likewise, we make it more difficult for adults to find and follow teens on Instagram by, among others, restricting adults from seeing teen accounts in “Suggested Users,” preventing them from discovering teen content in Reels or Explore, and automatically hiding their comments on public posts by teens; (6) we have developed technology to detect and remove bullying content even before it is reported, including, features such as bulk blocking, “tag” and “mentions” controls, blocking, “mute” interactions, to hide posts from certain accounts appearing on their Feed, without having to unfollow the account, “restrict” feature to “put some space” between themselves and another person’s account, hiding the person’s comments and messages on Direct (the private messaging feature on the Instagram service).

This approach helps to illustrate the suite of measures that we take against harmful content across the board with additional, specific measures or tools being appropriate in certain categories. We would also note that under the DSA, VSPS which are designated as VLOPs, will be required to conduct annual systemic risk assessments and to adopt appropriate and effective risk mitigation measures. This should therefore be taken into account in any risk mitigation requirements imposed under the Code particularly as it pertains to the dissemination of illegal content and the protection of minors. The approach to risk mitigation under the Code should also be principles based so as to ensure it is sufficiently adaptable and flexible to respond to changing and developing harms.

Question 3: Do you have reports, academic studies or other relevant independent research that would support your views? If you do, please share them with us with links to relevant reports, studies or research.

In order to assist an Coimisiún, we have set out below some documentation which we consider would be relevant and useful to the development of the first Code for VSPS:

- [Digital Trust and Safety Partnership Best practice Framework](#)
- [DTSP first report](#)
- [Netsafe Code](#) - CPOSH
- [Netsafe First report](#)

- **Meta reports** ([CSER](#), [Human Rights Impact Report](#), [Responsible Business Practices report](#))
- [Ofcom research](#)

As outlined in our response to **Question 1**, we would encourage An Coimisiún to draw upon the large body of research conducted by other regulators globally, for example, Netsafe in New Zealand or Ofcom in the UK. In particular, we would suggest drawing upon the CPOSH. We believe the CPOSH would assist An Coimisiún in the development of the Code as these principles have been established in order to promote safety while respecting the freedom of expression and other fundamental rights of users. They also recognise the transnational nature of the internet and take a systems-based approach to best practice standards.

SECTION 4 - WHAT ONLINE HARMS SHOULD THE CODE ADDRESS

Question 4: What approach do you think we should take to the level of detail in the Code? What role could non-binding guidance play in supplementing the Code?

What approach do you think we should take to the level of detail in the Code?

As noted above in response to **Question 1**, we believe that, in order for the Code to be most effective, it should be principles based and should establish a “*baseline*” of compliance measures for VSPS and a method of monitoring the effectiveness of that baseline. Additionally, the Code should align with established regulatory regimes (such as the DSA and the Terrorist Content Online Regulation (the TCO Regulation) - please see the response to Question 6 for further detail).

What role could non-binding guidance play in supplementing the Code?

We believe that, at this point, it is too early to say what role non-binding guidance could play in supplementing the Code. However, the AVMSD explicitly encourages the use of self- and co-regulation to meet its objectives. We would encourage An Coimisiún to make full use of self- and co-regulatory solutions in circumstances where the inclusion of an element of the Directive is not warranted in an online safety code.

If an Coimisiún were to consider, at a later stage, that non-binding guidance is required to supplement the Code, it is also worth bearing in mind that under Article 46 of the DSA, the European Commission (EC) may also draw up additional voluntary codes of conduct to tackle different types of systemic risks and illegal content.

Question 5: What do you think would be the most effective structure for the Code? What are the most important factors we should consider when we decide how to structure the Code?

As noted above in response to **Question 1**, we believe the Code will be most effective if it is principles based. Accordingly, the Code should generally identify the types of harmful content that

[REDACTED]

the Code is seeking to address and should set out the obligations applicable to VSPS, in general terms. To this extent, we highlight that the Code should not prescribe obligations in light of categories of content, as, often, the same types of measures are appropriate to deploy across different harm types.

We consider that the Code should adopt a two-tier structure setting out: (i) the baseline principles to be applied under the Code, regardless of the type of content/harm in question; and (ii) the supplemental measures VSPS may implement, based on the nature of their service(s). Indeed, Article 28b(3) AVMSD recognises that a range of complex factors need to be taken into account in determining whether measures are appropriate, including: a) the size and nature of the video-sharing platform service; b) the nature of the material in question; c) the harm the material in question may cause; d) the characteristics of the category of persons to be protected (for example, under-18s); e) the rights and legitimate interests at stake, including those of the person providing the video-sharing platform service and the persons having created or uploaded the material, as well as the general public interest.

These baseline measures, could for example, include the duty to put in place clear and easily accessible policies; the duty to have clear user reporting and appeal/complaint systems; the duty to effectively enforce those policies; the duty to take measures against users who are persistently abusing those policies. Supplemental measures that VSPS may implement but will not be obligated to do so in every case, could relate to technical measures, and safety tools which could be specifically tailored in view of the factors mentioned in Article 28b(3) AVMSD. This would enable VSPS to mitigate harms more effectively, as they would be able to adopt different mitigation measures in accordance with those factors mentioned above. This would also be consistent with the risk-based approach to regulation adopted under the DSA.

Question 6: How should we design the Code to minimise the potential for conflict and maximise the potential for synergies in how platforms comply with it and the DSA?

We strongly agree with this sentiment and the need to ensure that the Code does not conflict with the DSA (nor any other EU regulatory regimes).

To this end, when drafting the Code, an Coimisiún should keep in mind the importance of the uniform application of the DSA's harmonised rules to "*put an end to fragmentation of the internal market*" and "*ensure legal certainty*" (see Recital 4 DSA) and that Member States not adopt national measures dealing with requirements addressing the dissemination of illegal content online, as this is expressly recognised as an area which should be "*fully*" harmonised under the DSA (see Recital 9 DSA). Accordingly, the rules of the DSA should apply in respect of issues that are not addressed or not fully addressed by other Union legal acts as well as issues on which those other legal acts leave Member States the possibility of adopting certain measures at national level (see Recital 10 DSA).

The DSA addresses many of the same issues as AVMSD/OSMR and an Coimisiún should recognise that the measures taken by VSPS to comply with the DSA will likely assist them in meeting some, if not most, of the requirements imposed under the Code. In this manner, the measures implemented

by VSPS to address online harms can be considered holistically (i.e. as forming part of compliance solutions to both the DSA and the AVMSD). This would ensure that no conflict arises between the Code and DSA, which would create confusion and unnecessary regulatory burdens through parallel and duplicative mechanisms being imposed on VSPS.

The Code should further recognise that the DSA purposefully takes a tiered approach to regulation and the obligations that are applicable to services under the DSA are carefully balanced in light of the size and nature of their platform. This means that not all VSPS regulated by the Code will owe the same level of DSA obligations. In recognition of this fact, where an Coimisiún intends to introduce measures under the Code that are already prescribed for a given service type under DSA, e.g. VLOPs, then additional or contradictory measures should not be required simply because not all VSPS are VLOPs. By way of example, VLOPs are subject to risk assessment obligations under the DSA, whereas all other in-scope intermediary services are not. Likewise, an Coimisiún should take into account that the EU legislature chose to exempt non-VLOPs from those obligations.

In practice, where the Code seeks to apply measures or requirements that are already provided for under the DSA, this should mean that obligations are framed in a wholly consistent way with the DSA and do not contradict or go beyond the requirements of the DSA (in accordance with Recital 10 DSA). By way of example:

(i) **Transparency reporting:** Extensive periodic transparency reporting is required under the DSA per Articles 15, 24, 42. VLOPs have to report data every 6 months (see response to **Questions 9 and 16**).

(ii) **Risk Assessments:** The DSA introduces an important accountability framework for intermediary services. Certain VSPS, which have been designated as VLOPs under the DSA, are required to undertake risk assessments and implement risk mitigation measures in accordance with Articles 34 and 35 of the DSA (see response to **Question 18**).

(iii) **Turnaround Times for illegal content:** This is already harmonised by DSA, which does not prescribe specific turnaround times for the removal of illegal content and instead provides that notices should be processed in a “timely” way. A specific response time would contradict this standard and would not account for the nuance in assessing cases with differing levels of complexity (see response to **Question 9**).

(iv) **Commercial Communications and Ads Transparency:** Article 26(2) of the DSA already requires all VSPS to provide users with the ability to declare whether the content they provide is or contains commercial communications and in respect of advertisements, the DSA requires online platforms to identify, in a clear, concise and unambiguous manner, that the information is an advertisement (including through prominent markings) (Article 26(1) DSA) (see response to **Questions 8 and 21**).

(v) **Terms and Conditions:** Article 14 DSA requires that content moderation practices be reflected in terms and conditions, including information on applicable policies, tools and procedures (see response to **Question 14**).

(vi) **Reporting functionality for illegal content:** Article 16 DSA prescribes some requirements for

reporting mechanisms for illegal content including that such mechanism be “easy to access” and “user friendly” (see response to **Question 9**).

(vii) **Complaint handling and out-of-court-dispute settlement:** Articles 17, 20 and 21 DSA provide that certain content moderation decisions made by online platforms should provide a notice to the affected user and provide an effective complaint mechanism. It also provides that an individual can complain to an out-of-court-dispute-settlement body who may issue a non-binding decision (see response to **Question 16**).

By the same token, there may be some “appropriate measures” which do not form part of the DSA’s fully harmonised scope, but which feature within AVMSD/OSMR, e.g. measures to protect minors from certain types of harmful content. In this way, some of the measures which are deployed for the purposes of DSA may be leveraged further. To this end, where measures are proposed that are in accordance with Article 28(b) of the AVMSD but which are not clearly prescribed by DSA, these requirements should be identified so that it is clear that those measures apply specifically for VSPS (video) content. For example, with regard to the AVMSD Art. 28b(3)(d) and (e) requirement for providers to introduce a flagging mechanism, the Code could utilise the same principles for the mechanism provided for under the DSA that allows users to flag/report content they consider to be illegal, to cover certain types of harmful content. This can be drafted in such a way as to acknowledge the fact that many VSPS will already have user flagging/reporting functionality available for content that violates their Terms and policies.

Question 7: To what extent, if at all, should the Code require VSPS providers to take measures to address content connected to video content?

The Code should be clear as regards the content it intends to regulate. It will be extremely important that the Code does not exceed its legal remit and we would note that Article 28b of the AVMSD requires VSPS to implement appropriate measures in respect of certain categories of video content made available on those services. It does not require VSPS providers to address non-video content on their services.

Additionally, the recitals to the DSA clearly provide that it is intended to fully harmonise online safety rules applicable to intermediary services in the EU save to the extent other Union laws regulate other aspects of intermediary services, including AVMSD. It follows that while AVMSD should regulate video sharing elements of intermediary services, all content of those services, including the non-video content aspects of those services will be subject to the requirements of the DSA.

SECTION 5 - MEASURES TO BE TAKEN BY VIDEO-SHARING PLATFORMS

Question 8: How should we ask VSPS providers to introduce a feature that allows users to declare when videos contain advertising or other type of commercial communications? Should the Code include specific requirements about the form in which the declaration should take? What current examples are there that you regard as best practice?

How should we ask VSPS providers to introduce a feature that allows users to declare when videos contain advertising or other type of commercial communications?

Article 26(2) of the DSA already requires all VSPS to provide users with the ability to declare whether the content they provide is or contains commercial communications. Accordingly, as noted in **Question 6**, per Recital 10 DSA, the Code should align with the DSA in this context. Specifically, it should be made clear that mechanisms which comply with Article 26 DSA also comply with the Code, to the extent that such requirement is included.

In this circumstance, as noted in response to **Question 1**, given that VSPS will vary in terms of their size and the nature of content they make available, it is crucial that the Code adopts a principles-based approach, which gives VSPS some flexibility in respect of the features they adopt. The AVMSD also recognises that each of the measures listed in Article 28b(3) may not be appropriate for all VSPS.

Should the Code include specific requirements about the form in which the declaration should take?

This question points to a prescriptive approach rather than the principles-based approach discussed under **Question 1**. Accordingly, we believe that the Code should not include specific requirements about the form in which such a declaration should take, but rather impose a general obligation on VSPS (as appropriate) to put in place this functionality, to the extent that they don't already do so pursuant to the DSA.

Indeed, adopting such a principles-based approach ensures that the Code is future-proofed, while also allowing it to complement (rather than cut-across): (i) the DSA; and (ii) the work of other bodies, such as the Advertising Standards Authority of Ireland (**ASAI**). In this regard, it should be noted that the ASAI and Competition and Consumer Protection Commission intend to publish regulatory guidance about branded content in particular later this year, which will need to be taken into consideration by an Coimisiún.

More generally, we also note from our experience that specific language requirements rarely work effectively across multiple jurisdictions on the basis that translations often don't fully align in practice.

Further, the European Audiovisual Observatory's publication on the mapping of national rules applicable to VSPS⁹ has also acknowledged that other jurisdictions have generally transposed verbatim the requirements of AVMSD relating to the declaration of advertising/commercial communications. We submit that an Coimisiún should generally follow the same approach.

⁹ Mapping of national rules applicable to video-sharing platforms: illegal and harmful content online, a report prepared by the European Audiovisual Observatory for the European Commission ([here](#)), October 2022.

What current examples are there that you regard as best practice?

MPIL has adopted various measures in the context of its DSA compliance which (in our view) constitute best practice, given the prescriptive nature of the relevant provisions and the intent of the DSA. We summarise some of these measures below:

- **Advertising transparency:** As part of our compliance with Article 26 of the DSA, users are able to see who is benefitting from the advertisement and/or the organisation that is paying for the advertisement. This information - alongside information around the parameters used for targeting - is also available in the Meta Ad Library for one year after it serves its last impression. Advertisements that relate to social issues, elections or politics, are available for 7 years after serving their last impression.
- **Branded content:** For other types of commercial content e.g. commercial communications or branded content, Meta provides transparency in the form of a label applied to the relevant content and to comply with Article 39(2) of the DSA such content is also displayed in a repository¹⁰.

Question 9: How should we ask VSPS providers to introduce and design a flagging mechanism in the Code? How can we ensure that VSPS providers introduce the mechanism in a user-friendly and transparent way? How should we ask VSP Providers to report the decisions they've made on content after it has been flagged? To what extent should we align the Code with similar provisions on flagging in the DSA?

Like AVMSD Art. 28b(3)(d) and (e), Article 16 of the DSA requires providers to put in place easy to access and user-friendly mechanisms to allow users to flag/report content they consider to be illegal and for the providers to notify users of their decision. The DSA also contains transparency obligations in this regard (see paragraphs (4) to (6) of Article 16 of the DSA).

Many providers, including Facebook and Instagram, already had these mechanisms in place for content or accounts which violated their policies and, to comply with Article 16 of the DSA, have also developed such flagging mechanisms for illegal content. While it was already possible to report content as unlawful on both Facebook and Instagram, we have made this reporting option even more user-friendly. We work with content designers and user experience experts to ensure that such mechanisms are accessible and easy to use and understand.

The obligations that arise under Article 16 DSA apply to all VSPS that qualify as hosting services under the DSA, not just those which are VLOPs, and are more detailed than Article 28b(3)(d) and (e) and, accordingly, per Recital 10 DSA, the Code should align with the DSA in the context of illegal content reporting.

Additionally, with regard to reporting on decisions made in relation to content more generally, the Code should take into account the requirements under Articles 15, 24 and 42 of the DSA which

¹⁰ https://www.facebook.com/ads/library/branded_content/?source=onboarding

include extensive transparency reporting requirements on different types of reports and actions taken by relevant services. See, in particular, Article 15(1)(c) and (d) DSA.

Further, we note that turnaround times for illegal content is already harmonised by DSA, which does not prescribe specific turnaround times for the removal of illegal content and instead provides that notices should be processed in a “timely” way. The suggestion that the Code could go beyond the DSA in this area and “*could specify metrics about the timing and accuracy of moderation decisions and actions in relation to particular categories of content*” would contradict the “timely” standard and would not account for the nuance in assessing cases with different levels of complexity, as well as the need for a balancing assessment regarding the rights of affected individuals with respect to each removal or disabling of content as specifically required under the DSA.

In this context, an analogy can be drawn with interpretation of the word ‘expeditiously’ under Article 6 DSA, which the legislature intentionally avoided defining, As noted by the European Commission in the Impact Assessment accompanying the DSA proposal: “*national courts interpret “expeditiously” on a case-by-case basis taking into account a number of factors such as: the completeness of the notice, the complexity of the assessment of the notice, the language of the notified content or of the notice, whether the notice has been transmitted by electronic means, the necessity for the hosting service provider to consult a public authority, the content provider, the notifier or a third party and the necessity, in the context of criminal investigations, for law enforcement authorities to assess the content or traffic to the content before action is taken*”. Given that courts themselves take a case-by-case approach, we believe that it would be inappropriate for a sectoral regulator to seek to specify timing by means of an online safety code.

In short, the Code should be fully aligned with the DSA in this regard.

Question 10: What requirements should the Code include about age verification and age assurance? What sort of content should be shown by default to users who are logged out or in private browsing mode and whose age cannot be verified or assured? What evidence is there about the effectiveness of age estimation techniques? What current practices do you regard as best practice? Where accounts are not age verified should default privacy settings be used, should content default to universal content and should contact by others be more limited?

As described below, we consider that a combination of different measures, such as neutral age registration, reporting, alerts and verification (where appropriate) to be best practice and consider that a focus on wholesale age verification up front or for specific content types (which are already subject to moderation) would be privacy invasive and may be more easily circumvented or subject to fraud which could be more harmful for users.

Without prejudice to the foregoing, we believe that an Coimisiún should take into account the Data Protection Commission’s Fundamentals for a Child Orientated Approach to Data Protection (the **Fundamentals**¹¹), which outlines various recommendations relating to age verification, which already

¹¹ <https://www.dataprotection.ie/en/dpc-guidance/fundamentals-child-oriented-approach-data-processing>

apply to many (if not all) VSPS. Likewise, an Coimisiún should take into consideration that the EC is preparing an EU Youth Code which may touch upon the topics discussed herein.

Accordingly, in order to avoid fragmented and potentially conflicting requirements, an Coimisiún should consider whether to address this topic at this time and, in the case it decides to do so, an Coimisiún should adopt a principles-based approach that is consistent with other EU age assurance efforts.

We set out below some details of the core practices which we have deployed in this regard:

(i) **Neutral registration screen.** A date of birth screen should be presented without a pre-populated date of birth to ensure that people are not encouraged to circumvent an appropriate minimum age policy. If the prospective user enters a date of birth which would result in an age of between 5 and 12 years old (by way of example for Facebook and Instagram in Ireland), a screen should serve a generic error message informing them that they cannot create an account.

(ii) **Automated tools to prevent registration.** It is best practice to design blocking access for a period of time after repeat attempts to include an underage date of birth.


(iii) **Reporting underage users.** We have found that encouraging and facilitating processes for easy reporting of underage users is a proportionate measure, since we can then proceed to further verification checks before such users can continue to use the service. This avoids the need to disproportionately ask for identification from all users.

(iv) **Disabling violating linked accounts.** For platforms with multiple services, we also consider it best practice to enable simultaneous disabling across services where a user has been flagged as under age.

(v) **Predictive technology.** Age assurance technology such as age modelling – i.e. a combination of predictive technology and human review – to estimate the age of users, such as whether someone is above or below 18 years to help them receive an age-appropriate experience.

(vi) **Default privacy settings:** extensive obligations already apply to VSPS through the Fundamentals. As above, all accounts on Facebook and Instagram go through various age assurance steps both pre and post account opening. New teen accounts are then subject to various privacy content default settings which impact interactions with others as well as the content which may be displayed.

Question 11: What requirements should the Code have in relation to content rating? What do you consider to be current best practice? What experiences have you had using content rating systems on platforms and do you think they have been effective? What steps could we ask VSPS to take to ensure content is rated accurately by users?



In our opinion, to the extent that many VSPS already have robust terms and conditions in place which prohibit a wide range of harmful content, content rating requirements will not be necessary and will likely be ineffective. By way of example, hate speech, bullying and sexual activity are not allowed on Facebook and Instagram, and we also already apply warning screens on graphic content that does not violate our policies for under 18 users. Additionally, like other platforms, we already offer alternatives to content rating for minor users on Facebook and Instagram, such as age appropriate experiences (this includes reporting and blocking tools, parental resources and supervision tools, tools that allow users to hide like counts, referrals to resources, and time and usage management tools).

As such, it is unclear what the purpose/effect of such requirements would be and how they would work in practice as many types of content envisaged are already prohibited. This would therefore appear to be more akin to an optional measure for platforms that do not have robust policies already in place, although it may be appropriate for certain services.

Additionally, we believe that such a requirement could potentially create inconsistent experience for users through the European Union. Indeed, as an Coimisiún acknowledges in the CFI, the classification framework used for movies varies slightly across the EU; however, the ratings that apply to a given movie can vary significantly from Member State to Member State. Requiring individual providers to prescribe their own content rating systems which would be utilised by users would invariably result in inconsistent and potentially misleading content rating systems with varying outcomes for users.

We also believe that it would be unworkable to ask VSPS to ensure content is rated accurately by users as this would require proactive monitoring of content and amount to a general monitoring obligation. Requiring users to rate content is open to serious abuse and inaccuracies, not to mention a large amount of discrepancy.

Question 12: What requirements should the Code have in relation to parental control features? How can we ensure that VSPS providers introduce the mechanism in a user-friendly and transparent way? Can you point to any existing example of best practice in this area? Should parental controls be ‘turned-on’ by default for accounts of minors or where age is not verified?

The United Nations Convention on the Rights of Children has recognised that children have participation rights, which includes a right to have a say in matters affecting their own lives. Accordingly, we believe that any requirements in this regard should take into consideration the need to balance such participation rights with the challenges presented by introducing parental controls, such as, the need to verify parents/guardians and for the parents/guardians themselves to operate responsibly (e.g by ensuring that their child only sees age-appropriate content).

In addition, the requirement for online platforms accessible to minors to put in place appropriate and proportionate measures to ensure a high level of privacy, safety, and security of minors on their service under Article 28 of the DSA, should also be taken into consideration.

Based on our experience, we believe that we have struck the right balance by taking an “age

appropriate” approach. To this end, we have developed, in consultation with experts, parents and teens, tools to help users, including teens, have a safer, more supportive and age-appropriate experience online, and to help parents and teens navigate social media together. As noted in response to Question 2, this includes reporting and blocking tools, parental resources and supervision tools, tools that allow users to hide like counts, referrals to resources, and time and usage management tools. These have been designed to strike the balance between bringing parents into their teens’ experience and encouraging offline conversations, while still respecting teens’ privacy and autonomy.

By way of example, the current set of supervision tools allows parents and guardians whose teens opt-in to or agree to use supervision to, inter alia, (i) view how much time their teen spends on the Instagram service across devices in the last 7 days; (ii) set daily time limits; (iii) get notified when their teen shares that they have reported someone; (iv) view and receive updates on what accounts their teen follows and the accounts that follow their teen; (v) see which accounts their teen is currently blocking; and be notified if their teen changes any of these settings.

Likewise, we have also made resources easily available for teens, and their parents and guardians, to ensure that they are fully informed of the applicable standards and available options, such as (1) educational resources with information about the privacy and safety tools available to their teens for parents which include Parents Portal¹², Parent Centre¹³ and Parent’s Guide¹⁴), (2) Family Centre¹⁵, a place for parents and guardians (with their teens’ permission) to oversee their teens’ accounts on Instagram and set up and use supervision tools, (3) Education Hub¹⁶, where parents and guardians can access resources from experts and review helpful articles, videos and tips on topics like how to talk to their teens about safe use of social media.

Question 13: What requirements should the Code contain to ensure that VSPS provide for effective media literacy measures and tools?

We acknowledge that media literacy is an important issue in terms of (amongst other things) enabling access to information and allowing users to create content in a responsible and safe manner. However, as is recognised by recital 59 of the AVMSD, VSPS are part of a broader ecosystem of stakeholders responsible for promoting the development of media literacy in all sections of society.¹⁷

It is important to note that the definition of media literacy in the Broadcasting Act 2009 is geared towards traditional broadcasting services and does not reflect the definition set out in Recital 59 of the 2018 AVMSD amending Directive: ‘Media literacy’ refers to skills, knowledge and understanding that allow citizens to use media effectively and safely. In order to enable citizens to access information and to use, critically assess and create media content responsibly and safely, citizens need to possess advanced media literacy skills. Media literacy should not be limited to learning about

¹² See <https://www.facebook.com/safety/parents>

¹³ See <https://about.instagram.com/community/parents>

¹⁴ See <https://about.instagram.com/community/parents/guide>

¹⁵ See <https://familycenter.meta.com>

¹⁶ See <https://familycenter.meta.com/education/>

¹⁷ Recital 59 of the AVMSD: “video-sharing platforms providers, in cooperation with all relevant stakeholders, promote the development of media literacy in all sections of society, for citizens of all ages, and for all media and that progress in that regard is followed closely”

tools and technologies, but should aim to equip citizens with the critical thinking skills required to exercise judgement, analyse complex realities and recognise the difference between opinion and fact. It is therefore necessary that both media service providers and video-sharing platforms providers, in cooperation with all relevant stakeholders, promote the development of media literacy in all sections of society, for citizens of all ages, and for all media and that progress in that regard is followed closely”.

Considering the nature of this ecosystem, and indeed the myriad of efforts that encompass media literacy, this is an area where, as noted in **Question 4** above, we would encourage an Coimisiún to make use of self- and co-regulatory solutions.

Question 14: How should we ask VSPS providers to address online harms in their terms and conditions in the Code, including the harms addressed under Article 28b? How should key aspects of terms and conditions be brought to users’ attention? What examples are there of best practice in relation to terms and conditions including content moderation policies and guidelines?


As noted in response to **Question 6**, Article 14 of the DSA already requires that all intermediary services have in place terms and conditions with “*clear, plain, intelligible, user-friendly and unambiguous language, and [that] shall be publicly available in an easily accessible and machine-readable format*”, which include information on any restrictions that they impose on the use of their services. In practice, this means that the DSA already requires that content moderation practices be reflected in our terms and conditions, including information on applicable policies and procedures.

Article 14 of the DSA is a comprehensive provision and is an example of maximum harmonisation under the DSA. As such, and as mentioned in response to **Question 6**, in accordance with Recital 10 DSA, the requirements under the Code should be framed in a wholly consistent way with the DSA and, to the extent necessary, the Code should mirror that provision. Additionally, it should be made clear that terms and conditions which comply with the DSA also comply with the Code

We believe that our approach to content policies (or Community Standards/Guidelines) and Terms of Service – which have been adapted in compliance with the DSA – reflect best practice in this area. We have over eighteen years of experience in developing and enforcing content policies across our services. We dedicate significant time and resources into developing content policies, and indeed maintaining such policies to reflect evolving trends in technologies, products, circumvention techniques used by bad actors and societal behaviours (see response to **Questions 2 and 15** for more detail on the development and enforcement of our content policies).

We take a range of steps to help make the terms and policies of our services easy to understand and accessible. We offer tools to help people make safe choices on our platforms and we work to be transparent about how we address these issues. That’s why we make our policies – including the Facebook Terms of Service and Community Standards, the Instagram Terms of Use and Community Guidelines, and other specific policies (such as our Commerce Policies and Ad Policies) – available online to everyone, via our Transparency Centre¹⁸.

¹⁸ <https://transparency.fb.com/en-gb/policies/>



As a result, our policies and Terms of Service are designed to be accessible – both through the relevant apps and websites –, user-friendly and carefully drafted to be easy to follow whilst providing users with an appropriate level of detail, and are made available in a range of languages, to make them easy to understand. Meta’s transparency centre also provides information on relevant practices in terms of how we enforce those policies and how we handle complaints from users in relation to our content enforcement decisions¹⁹.

In practical terms, users are presented with the applicable terms of service when they sign-up to Facebook or Instagram – either through the relevant apps or websites –, and are also directed to the applicable terms when we enforce our policies. For example, when we enforce those policies against a user’s account or content, we provide them with relevant information so that they can understand why we took action.

We also release a quarterly Community Standards Enforcement Report (**CSER**), which shows how we are doing at enforcing our policies. This kind of transparency lets people see clearly how we are addressing safety issues and helps us get much-needed feedback.

Question 15: How should we ask VSPS providers to address content moderation in the Code? Are there any current practices which you consider to be best practice? How should we address automated content detection and moderation in the Code?

At the outset, we believe that an Coimisiún should not prescribe requirements for content moderation. Content moderation is a complex, evolving and multi-faceted approach that necessarily varies in detail across different services. As we have provided in **Question 1**, a principles-based approach in the Code would ensure that VSPS take steps to develop appropriate terms and conditions to prohibit certain types of harmful content and should effectively enforce those policies. The requirement to undertake a systemic risk assessment and to publish data in relation to content moderation decisions, as provided for in the DSA, means that there is accountability and transparency surrounding the enforcement of those terms and conditions. This systems based approach strikes an appropriate balance in our view.

Article 15 eCommerce Directive

We agree that Article 15 of the E-Commerce Directive precludes the imposition of any general monitoring obligation on VSPS providers, and we note that Article 8 of the Digital Services Act also adopts the same approach. We respectfully note that the imposition of any obligations to monitor particular categories of content would therefore be precluded by these provisions of the E-Commerce Directive and the Digital Services Act.

In the *Glawischnig* case (Case C-18/18), the Court of Justice of the European Union (“CJEU”) held that Member States can only impose monitoring obligations on intermediaries (such as VSPS providers) in relation to “*a specific case*” – not a whole category. In particular, the CJEU held that where a specific item of content has already been declared unlawful by a competent court, a Member State can require intermediaries to remove content that is identical to that specific item of unlawful content,

¹⁹ <https://transparency.fb.com/enforcement/>

[REDACTED]

or content that is so similar to the specific item of unlawful content that it would not require any independent assessment by the intermediary. However, an obligation to monitor entire categories of content would clearly not be in “*a specific case*”, nor would it relate to content already declared unlawful by a competent court. Moreover, it would clearly require VSPS providers to conduct independent assessments of content in order to determine which content falls within any such categories. Imposing any such monitoring obligations on VSPS providers would therefore clearly be contrary to the CJEU’s ruling.

There are strong reasons behind the CJEU’s ruling and jurisprudence in this area. As Advocate General Øe pointed out in Joined Case C-682/18 and C-683/18, requiring VSPS providers to extend any monitoring beyond “*specific cases*” (for example, to remove content on the grounds that such content falls within a certain category) would inevitably lead to ‘over-removal’ by some service providers, who might understandably seek to reduce the risk of liability in borderline cases. Such ‘over-removal’ would, as the Advocate General put it, “*pose an obvious problem in terms of freedom of expression*”.

Question 16: What requirements should the Code include about procedures for complaint-handling and resolution, including out-of-court redress or alternative-dispute resolution processes? To what extent should these requirements align with similar requirements in the DSA? What current practices could be regarded as best practice? How frequently should VSPS providers be obliged to report to an Coimisiún on their complaint handling systems and what should those reports contain? Should there be a maximum time-period for VSPS providers to handle user complaints and if so, what should that period be?

As an Coimisiún has noted in the CFI, Articles 17 and 20 of the DSA already require that certain content moderation decisions made by platforms should provide a notice to the affected user and provide an effective complaint mechanism, and Article 21 entitles users to resolve disputes in relation to complaints via a certified out-of-court dispute settlement body, who may issue a non-binding decisions.

Likewise, intermediary service providers are also subject to certain transparency reporting requirements under Articles 15, 24 and 42 of the DSA (as appropriate), including a requirement to prepare transparency reports on the number of complaints received through internal complaint-handling systems. Notwithstanding the introduction of these requirements under the DSA, we have for many years published periodic transparency reports on our content moderation efforts and have worked to expand on this over time. We also publish data externally on a recurring basis on our response to government takedown requests and data requests. These efforts have been built upon to meet the DSA’s additional transparency requirements.

Additionally, while we appreciate that an Coimisiún wishes to hold VSPS accountable, in our view, prescriptive turn-around-times create the wrong incentives by overlooking the challenges of nuanced legal review, e.g. balancing freedom of expression, privacy rights and safety. We would encourage An Coimisiún to require VSPS to handle user complaints in an efficient and timely manner, thus ensuring that complaints are effectively dealt with, without imposing prescriptive turn-around times. This would also be in line with the requirements under the DSA in respect of the manner in which hosting

services are required to deal with notices submitted through the notice and action mechanism (see response to **Question 9**).

At Meta, we have spent considerable time and resources developing a system which we believe strikes an appropriate balance. Our systems prioritise harmful content with the most views, which allows us to quickly remove content that is having the greatest effect on our users (e.g. terror content or content child abuse before looking at more harmless content types).²⁰ We believe that an Coimisiún should consider a similarly flexible approach which allows VSPS to deal with the most dangerous and harmful content first.

The DSA provisions in this regard are sufficiently detailed and are an example of maximum harmonisation under the DSA. As such, and as mentioned in response to **Question 6**, in accordance with Recital 10 DSA, the requirements under the Code should be framed in a wholly consistent way with the DSA and, to the extent necessary, the Code should mirror such provisions. There is a significant risk of confusion and conflict if an Coimisiún chooses not to do so (e.g. in the event of parallel complaints being raised under the complaint handling systems for each regime).

Question 17: What approach do you think the Code should take to ensuring that the safety measures we ask VSPS providers to take are accessible to people with disabilities?

Noting that the legislative accessibility requirements in Article 7 of the AVMSD apply only to media service providers and not VSPS, we would strongly encourage an Coimisiún to be aware that there are other specific considerations in this regard under EU Law. In particular, the European Accessibility Act (EAA) introduces accessibility requirements for certain services. The European Commission is also required to encourage and facilitate the drawing up of codes of conduct for accessibility at a European Union level under Article 47 of the DSA. In light of these considerations, we believe it may be premature to deal with accessibility in the Code given the possibility of it being in conflict with or superseded by this EU-wide code.

Question 18: What approach do you think the Code should take to risk assessments and safety by design? Are there any examples you can point us towards which you consider to be best practice?

The DSA introduces an important accountability framework for intermediary services. Certain VSPS, which have been designated as VLOPs under the DSA, are required to conduct annual systemic risk assessments and to adopt appropriate and effective mitigation measures in light of the findings of the risk assessment (Articles 34 and 35 of the DSA). VLOPS will also be required to provide reports on the risk assessments to relevant supervisory authorities, in our case, to the EC and An Coimisiún (as Meta's Digital Services Coordinator). Such reports will also be made publicly available (albeit at a later date) (see article 42(4)a of the DSA).

The risk assessment requirements under the DSA are significant and extensive and therefore, as mentioned in response to **Question 6**, in accordance with Recital 10 DSA, any risk assessment and

²⁰ Meta, Transparency Centre, Prevalence ([link](#)).

mitigation requirements under the Code should be framed in a wholly consistent way with the DSA and, to the extent necessary, the Code should mirror the DSA provisions. By way of example, VLOPs are subject to risk assessment obligations under the DSA, whereas all other in-scope intermediary services are not. Likewise, an Coimisiún should take into account that the EU legislature chose to exempt non-VLOPs from those obligations and, thus, an Coimisiún should not impose similar obligations on non-VLOPs.

To the extent that an Coimisiún decides to include risk assessment requirements within the Code, it is imperative that these take full account of the risk assessment obligations that some VSPs are already subject to under the DSA. The Code should therefore only require VSPs which are also designated as VLOPs to assess risks which would not already be covered by the DSA risk assessment obligations.

The Code should also be principles based so as to ensure it is sufficiently adaptable and flexible to respond to changing and developing harms.

Additionally, it should be made clear that risk assessments and mitigations that comply with the DSA also meet any risk mitigation requirement in the Code.

Question 19: How do you think that cooperation with other regulators and bodies can help us to implement the Code for VSPs?

As already mentioned (please see our responses to **Questions 1** and **6** for more detail), we believe that a harmonised approach to regulation is more effective and efficient. As such, cooperation with the European Commission and other Digital Services Co-ordinators will be essential. We also believe that it would be beneficial for an Coimisiún to cooperate with the UK's Ofcom which also oversees a comparable (though not identical) VSP regime.

To the extent that the Code touches on issues which may overlap with our obligations under data protection legislation, (for example, in respect of the protection of minors), cooperation with relevant data protection authorities including the Data Protection Commission (**DPC**) in Ireland is needed to ensure consistency. By way of a specific example, an Coimisiún should consider the contents of the Fundamentals. Meta is careful to comply with the Fundamentals and hence it would be unnecessary to cover in a Code anything which already forms a requirement of the Fundamentals.

Additionally, cooperation with other Irish sectoral regulators or bodies will likely be required and/or appropriate: For example the Competition and Consumer Protection Commission or the Advertising Standards Authority of Ireland. As such, we believe that an Coimisiún should involve these regulatory bodies in developing relevant codes in consultation with industry.

Question 20: What approach do you think we should take in the Code to address feeds which cause harm because of the aggregate impact of the content they provide access to? Are there current practices which you consider to be best practice in this regard?

As an overarching point, any requirement built into the Code should be based on assumptions which

are backed by evidence and rooted in research.

It should be noted that the AVMSD is silent on this but it is addressed by the DSA (see, in particular, Articles 34(2)(a), 35(1)(a) and (d) and 38 which all apply to VLOPs. Accordingly, no additional obligation in this regard should be placed on VLOPs. For all other VSPS, an Coimisiún should take into account that the EU legislature chose to exempt non-VLOPs from those obligations.

Notwithstanding the above, MPIL is confident that it has pioneered many best practices for the benefit of users (and in particular, younger users). For example, we refer an Coimisiún to MPIL's youth safety strategy²¹. It is also worth noting that MPIL, along with all other VLOPs, is required (under Articles 27 and 38 of the DSA) to provide transparency and introduce at least one option for each of its recommender systems which is not based on profiling. We have launched 'System Cards'²² - which have been added to Meta's Transparency Center and cover FB and IG Feed, Stories, Reels and other surfaces - which give information about (a) how our AI systems rank content, (b) some of the predictions each system makes to determine what content might be most relevant to users, as well as (c) the controls users can use to help customise their experience. For each of the recommender systems that have 'System Cards' in the Transparency Center, a Facebook or Instagram user can access features or experiences that allow for a non-personalized experience. We consider that this will further address any potential negative effects that an Coimisiún considers might arise from the aggregate impact of content.

In addition, we have rolled out a number of well-being features on Instagram, such as, sensitive content control; daily limit; mute push notification (aka pause all); take a break for reels; quiet mode; and alternative topic nudge.


Question 21: Do you have any views on how requirements for commercial content arranged by a VSPS provider itself should be reflected in the Code?

Article 9(1) of the AVMSD is clear on the requirements that should be imposed on VSPS with respect to audiovisual commercial communications that are marketed, sold or arranged by and, accordingly, we consider that the Code should be limited to transposing these requirements of the AVMSD into Irish law, by, for instance, requiring that said requirements be reflected in VSPS terms and policies.

By way of example, Meta has strict advertising policies for advertising to all users, which impose high standards on paid advertising. Among other things, the Advertising Policies (applicable to Facebook and Instagram) strictly prohibit ads promoting the sale or use of certain types of products for all users, such as tobacco and related products, drugs and drug-related products, and adult content. Meta further age-restricts (i.e., 18+) ads for certain products or services, like alcohol, dating services, gambling, sexual and reproductive health products, dating services, and weight loss products. Our policies also provide for restrictions on the personalisation of advertising to minors by default, meaning that age and general location are the only information about a minor that Meta uses to show them ads, ensuring that teens see ads that are meant for their age and products and services available where they live.

²¹ See <https://about.meta.com/actions/safety/audiences/youth/>

²² See <https://transparency.fb.com/features/explaining-ranking>



How those policies are enforced may also be taken into consideration by an Coimisiún in the Code. For instance, our ad review process starts automatically before ads begin running, and is typically completed within 24 hours, although it may take longer in some cases. If a violation is found at any point in the review process, the ad will be rejected. We use automated and, in some instances, manual review to enforce our policies and, beyond reviewing individual ads, we also monitor and investigate advertiser behaviour, and may restrict advertiser accounts that don't follow our advertising policies, Community Standards or other Meta policies and terms.

In any case, consideration should be taken to the requirements already prescribed in Article 26 of the DSA (see response to **Question 6**).

Question 22: What compliance monitoring and reporting arrangements should we include in the Code?

As noted in response to **Question 18**, the DSA introduces an important and extensive accountability framework for intermediary services which will also apply to VSPS to differing degrees i.e. depending on the type of “intermediary service”. Accordingly, when determining the appropriate requirements for compliance monitoring and reporting arrangements in the Code, we would ask that an Coimisiún bear in mind the considerable reporting obligations which already exist notably under the DSA, and to the greatest extent possible, avoid the creation of unnecessary or duplicative obligations:

(i) Per Articles 15, 24 and 42 of the DSA, extensive periodic transparency reporting is required (see response to **Questions 9 and 16**).


(ii) Per Article 34 and 35 of the DSA, VLOPs are required to conduct annual systemic risk assessments and to adopt appropriate and effective mitigation measures in light of the findings of the risk assessment and to prepare reports on the risk assessments and mitigation measures which are to be provided to relevant supervisory authorities (see response to **Question 18**).

(iii) Per Article 37 of the DSA, VLOPs shall be subject, at least once a year, to independent audits to assess compliance with a broad range of requirements set forth in the DSA (i.e., Articles 11 to 48 of the DSA), a report of which shall be provided to the relevant supervisory authorities.

Under Article 42(4)a of the DSA, the risk assessment and mitigation report and the audit report have to be provided to an Coimisiún, as Meta’s Digital Services Coordinator.

Such assessments and reporting requirements under the DSA are significant and extensive and, therefore, as mentioned in response to **Question 6**, compliance with these obligations should be regarded as part of the AVMSD/OSMR compliance solutions to the extent that they achieve similar objectives.

Additionally, it is worth taking note of Article 41(6) of the DSA and the role envisaged for the management body of VLOPs in reviewing and approving strategies in relation to risk management and mitigation.



As an Coimisiún can appreciate, in an organisation as large as Meta, the production of reports of this nature takes a considerable amount of time as their compilation involves multiple stakeholders. All data which is published goes through a rigorous checking process and multiple tiers of review. We respectfully request that before determining any reporting obligations under the Code, you consider the possibility that the data which you may need to perform your functions may already be available via other channels and, in the event that additional measurements are required, to allow for sufficient synergies between regulatory reporting and validation windows e.g. when reports should fall due, what time period they should cover and how long services should be given to validate the relevant data.

Question 23: Should the Code have a transition period or transition periods for specific issues? Which areas touched on in this Call for Inputs may VSPS providers require time to transition the most? What time frame would be reasonable for a transition period?

We believe that the Code should have at the very least a six-month implementation period.

However, if the Code takes a prescriptive approach to requirements (which we believe would be contrary to the objectives of the AVMSD), then longer transition periods should be provided for. The duration of such long transition periods would be determined by the specific level of change required by the relevant Code requirements.

VSPS Regulation
Coimisiún na Meán
2-5 Warrington Place
D02XP29
Ireland

04 September 2023

Ofcom submission to 'Call for Inputs: Online Safety'

Ofcom is the United Kingdom's (UK) communications regulator, overseeing sectors including fixed-line and mobile telecoms, the airwaves on which wireless devices operate, post, and TV and radio broadcasting. We also regulate online video services established in the UK, including on-demand programme services, and video-sharing platforms (VSPs). We are currently preparing to regulate online safety.

We welcome the opportunity to provide input into the development of Coimisiún na Meán's first binding online safety code. The services that we regulate, as well as the safety risks that we seek to protect individuals from, are global in nature. We see an important opportunity for regulators across countries and regions to share experience, expertise, and evidence as we collectively drive improvements in online safety.

Regulatory coordination and cooperation benefits us as regulators and helps us further our respective domestic objectives. It is also helpful for the services that we regulate, and coordination around regulatory expectations and supervisory approaches can promote services' compliance across jurisdictions.

In what follows, we will provide some insights into how we have approached the regulation and supervision of video-sharing platforms that have notified themselves to Ofcom. We will outline our regulatory strategy – what we do and what we aim to achieve – as well as the learnings we have gathered from supervising a regulatory regime that has been in operation for over two years.

We hope that our approach and experiences can provide useful insights for Coimisiún na Meán as it develops its upcoming code.

The UK video-sharing platform framework

The UK VSP Framework is set out in Part 4B of the Communications Act 2003 (the Act) and derives from the provisions of the revised EU Audiovisual Media Services Directive (AVMSD) 2018. The requirements for platforms came into force on 1 Nov 2020. As a transposition of EU law, our approach to VSP regulation therefore starts from the same place as the Irish framework.

Duties on VSPs

The intent of the VSP Framework is to protect people from harmful video content. It therefore places a duty on VSP providers to take and implement appropriate measures to protect the general public

from ‘relevant harmful material’^{1 2}. They must also protect children who are under-18 from ‘restricted material’^{3 4}. Ofcom refers to ‘relevant harmful material’ and ‘restricted material’ together as ‘harmful material’.

Schedule 15A of the VSP legislation lists some measures that VSP providers must take, if appropriate for their platform, to fulfil their duties to protect users from harmful material. Where providers take Schedule 15A measures to protect users from harmful material, they are required to implement them effectively, and in a way that achieves the protection for which the measures are intended⁵.

The measures concern:

- terms and conditions relating to harmful material;
- flagging, reporting, or rating mechanisms;
- appropriate access control measures to protect under-18s, such as age assurance systems and/or parental control measures in relation to restricted material;
- easy-to-use complaints processes; and
- media literacy tools and information.

VSP providers are required to determine which of the Schedule 15A measures are appropriate for their platform, based on whether it is practicable and proportionate for that provider to implement it, considering factors including:

- the size and nature of its platform;
- the type of material on the platform and the harm it might cause;
- the characteristics of users to be protected;
- the rights and legitimate interests of users, the general public and the provider; and,
- any other non-Schedule 15A measures already implemented on the platform.⁶

Further to the above measures, the VSP Framework also seeks to ensure that certain standards are met when advertising controlled by a service provider is delivered on VSPs.

To be subject to the VSP Framework in the UK, services must meet the definition of a VSP and be established in the UK⁷. Ofcom’s guidance, [‘Video-sharing platforms: who needs to notify to Ofcom?’](#) aims to help services understand whether they are subject to the Framework. Importantly, the VSP Framework is based on self-notification to Ofcom. It is the responsibility of the *service* to determine

¹ Schedule 15A of the VSP Framework transposes Article 28b of the EU AVMS directive, setting out a list of measures which might be appropriate for service to take to ensure the required protections.

² ‘Relevant harmful material’ is video content which would be considered a criminal offence under laws relating to terrorism; child sexual abuse material; and racism and xenophobia.

³ Communications Act 2003, Section 368Z1, Available at: <https://www.legislation.gov.uk/ukpga/2003/21/section/368Z1>

⁴ ‘Restricted material’ is video content which has or would be likely given an R18 certificate, or video content not suitable for British Board of Film Classification (BBFC) classification, or material that might impair the physical, mental, or moral development of under-18s.

⁵ Communications Act 2003, Section 368Z1 (2), Available at: <https://www.legislation.gov.uk/ukpga/2003/21/section/368Z1>

⁶ For more information on the proportionality and practicality criteria for measures, see Communications Act 2003, Section 368Z1 (4), Available at: <https://www.legislation.gov.uk/ukpga/2003/21/section/368Z1>

⁷ Communications Act 2003, Section 368S, Available at: <https://www.legislation.gov.uk/ukpga/2003/21/section/368S>

whether it meets the statutory criteria for the Framework and whether it needs to notify Ofcom⁸. At this point in time, 20 services are notified under the Framework.

Ofcom's duties

Ofcom has been designated as the responsible UK regulator for the VSP Framework. As such, Ofcom is required to take such steps as necessary to secure compliance by VSP providers with their obligations under Part 4B of the Act. Ofcom is also required to draw up and publish guidance concerning the measures in Schedule 15A which might be appropriate for VSP providers to take to protect users from harmful material, and the implementation of such measures. Our guidance on these matters can be read [here](#).

Ofcom has extensive information-gathering powers to enable it to fulfil its duties. These powers can be used, among other things, for assessing and monitoring compliance by VSP providers, conducting investigations into suspected contraventions of the VSP requirements, and gathering the information needed to produce and publish reports under section 3668Z1(1) of the Act. These reports, also known as transparency reports, focus on measures taken by VSP providers and the steps taken to implement those measures effectively.

Where, following an investigation, Ofcom determines that a VSP provider has failed to comply with the Act, it has the power to issue enforcement notifications (which might set out the steps required to remedy a contravention) and/or impose financial penalties of up to £250,000 or 5% of qualifying revenue, whichever is greater. In certain circumstances, Ofcom may also suspend and/or restrict a service⁹. Ofcom's enforcement of the VSP Framework is undertaken in accordance with Ofcom's [Enforcement Guidelines](#). Annex 4 of the Enforcement Guidelines provides information on how they apply for the VSP Framework specifically.

Ofcom is a public body and as such, when issuing guidance to help services comply with the VSP Framework and when pursuing investigations and enforcement actions, its activities are undertaken in light of the Human Rights Act 1998 and the European Convention on Human Rights.

Relationship between the VSP Framework and the future Online Safety Act

We expect that all services currently in scope of the VSP Framework will also be in scope of the Online Safety Bill (OSB), when enacted¹⁰.

When the OSB comes into force, all pre-existing UK-established VSPs will enter a transition period. During this period, they will be exempt from having to comply with most duties under the OSB and will continue to be regulated under the existing VSP Framework. The date at which the transition period ends – and when the VSP Framework is repealed – will be specified in secondary legislation to be made by the UK Government.

Following the transition period, pre-existing, UK-established VSPs will be regulated under the OSB and all duties will apply in full. Further information on this process can be found [here](#).

⁸ Failure to provide advanced notification constitutes a breach of the Framework, which is enforceable by Ofcom.

⁹ Communications Act 2003, Section 368Z3, Available at: <https://www.legislation.gov.uk/ukpga/2003/21/section/368Z3>

¹⁰ More details on our preparations for this new framework can be found [here](#).

Ofcom's regulatory strategy

[Ofcom's approach to VSP regulation](#) has four broad **aims**, which shape our work and underpin how we drive good user outcomes:

- Raise standards in user protections;
- Rapidly identify and address areas of non-compliance;
- Increase transparency across the industry; and,
- Get industry and ourselves ready for the OSB, when enacted.

We set our VSP **strategic priorities** against these aims, and we pursue them in line with the principles of proportionality, flexibility, and dynamic supervision. We discuss our strategic priorities in more detail in Section Four.

Regulating with proportionality and flexibility

Ofcom is not responsible for regulating *all* VSPs, as our duties apply only to services which meet the UK jurisdictional criteria. For all other VSPs active in the UK, we rely on our regulatory counterparts in other European countries – including Ireland – to achieve the desired user safety outcomes. At the time of writing, 20 service providers, including TikTok, Snap, Twitch, and OnlyFans, are currently notified to Ofcom as meeting the relevant criteria for being regulated in the UK. These 20 services offer a range of different experiences and vary significantly in terms of size, reach and resources.

The diversity of the VSPs in scope means that the risks they might pose, and the protections they need to offer to keep users safe, both vary. The VSP legislation – and the guidance that Ofcom develops on the basis of it – seeks to reflect this diversity. It identifies a range of protective measures but does not prescribe which ones a platform should take, nor does it require a uniform approach across platforms. The focus is on ensuring that the measures that platforms *do* take are appropriate to protect their users from harmful material. Moreover, the onus is on VSP providers to determine how best to manage the risks that their services pose and to take action that is proportionate to the risk of harm and tailored to the circumstances they face.

Ofcom expects VSP providers to be proactive, anticipating risks and taking proportionate preventative steps. This does not mean they are expected to undertake general monitoring for harmful content. It means providers should regularly and systematically work to combat existing and emerging risks, by having effective protection measures in place that take account of Ofcom's [guidance on appropriate measures](#).

Our guidance is not prescriptive about the specific approach or technical tools that platforms should adopt. Rather, it outlines good practice that services should consider when implementing the legislative measures. We also encourage providers to use relevant evidence to make decisions about their measures and to collect data on their effectiveness.

A dynamic approach to supervision

The VSP Framework covers a broad range of online video content and services and places considerable responsibility on, and affords considerable discretion to, VSPs to determine how to achieve the desired regulatory outcomes. Ofcom has a critical role to play in supporting providers in navigating the regulatory objectives and ensuring that they take appropriate measures to protect their users.

In supervising the VSP Framework, Ofcom aims to be rigorous but fair, having regard to our regulatory principles of transparency, accountability, proportionality, and consistency. We expect platforms to engage constructively and openly with Ofcom and be willing to make improvements to enhance the effectiveness of the safety measures they deploy. In our engagement with notified providers, we discuss their processes and responses to specific issues and seek to better understand the

effectiveness and limitations of their measures. Where we have concerns about a VSP provider's measures or about safety risks on the service, we will generally seek to resolve these concerns informally and by agreeing a roadmap with the service, as per our Enforcement Guidelines. This constructive and deliberative approach often provides the quickest and most efficient route to improving VSP measures and ultimately ensuring users are better protected.

Ofcom takes an evidence-based approach to supervision and assessment, leveraging our information-gathering powers and data science capabilities to target the greatest risks of harm. Ofcom does not have a role in responding to or adjudicating on individual user complaints, but we do monitor trends in [complaints made directly to us](#) to help us identify where there might be issues with providers' protection measures or new harms arising. We also collect information and insights from regular engagement with a broad range of actors like tech safety groups, civil society organisations, and charities with an interest in online safety, and we actively track user experiences online and monitor trends through our extensive programme of research¹¹.

With respect to potential breaches of the Framework, our first approach is to work, where possible, with stakeholders to solve potential breaches informally. We do however have robust powers to take formal enforcement action in cases where, for example, it appears unlikely that engagement will achieve the required improvements. If, following an initial assessment, Ofcom decides that formal enforcement action is necessary, we will investigate the issue to determine if there has been a breach and what further action might be appropriate¹².

Ofcom has both completed and is currently undertaking a number of investigations and enforcement actions under the VSP Framework. In March 2023 Ofcom fined Tapnet Ltd – which provides the video-sharing platform RevealMe – £2,000 after the company did not respond to a statutory request for information¹³. In May, Ofcom opened an own-initiative investigation into Secure Live Media Ltd (SLM), in respect of the video-sharing platform service CamSoda and SLM's compliance with its statutory obligations under Part 4B of the Act¹⁴. And as a final example, Ofcom recently opened an enforcement programme on age assurance across the adult VSP sector. Enforcement programmes seek to examine a problem or concern that relates to a particular group of stakeholders, or to a whole sector. The programme was opened after our finding that many notified adult VSPs do not appear to have measures that are robust enough to stop children accessing pornographic material¹⁵.

Our experience to date

Lessons from year one

The VSP Framework came into effect in November 2020. In the initial period Ofcom developed and published a range of preparatory materials. Ofcom began its supervisory activities in October 2021 and at that time published our guidance for VSPs on implementing measures to protect users.

At the end of our first full year of regulating VSP in October 2022, [we published a report](#) outlining how notified VSPs had performed against our regulatory expectations as well as the learnings we had garnered through executing this new form of regulation. Much of the report's learning were informed by the information requests that we had sent to services as well as ongoing supervisory engagement in Year One.

¹¹ See the annex for a full list of relevant research publications.

¹² The various enforcement powers at our disposal are outlined in Section Two.

¹³ More information on the Tapnet investigation can be found [here](#).

¹⁴ More information on the SLM investigation can be found [here](#).

¹⁵ More information on the ongoing enforcement programme can be found [here](#).

The [report](#) includes several high-level insights that are likely to be of interest and relevance to other VSP regulators. Through our regulation of notified VSPs in 2021-2022 we found that:

- **All platforms have safety measures in place**, including rules on what kinds of video material is allowed. Some platforms made changes to their measures in direct response to being regulated under the VSP Framework.
- **Platforms generally provided limited evidence on how well their safety measures are operating to protect users**. This creates difficulty in determining with any certainty whether VSPs' safety measures are working consistently and effectively.
- **More robust measures are needed to prevent children accessing pornography**. Some adult VSPs' access control measures are not sufficiently robust in stopping children accessing pornography. This learning informed our decision to open an enforcement programme into the adult VSP sector earlier this year (as referenced in Section Three above).
- **Some platforms could be better equipped for regulation**. Some platforms are not sufficiently prepared and resourced for regulation. Going forward, we will be looking for platforms to improve and provide more comprehensive responses to Ofcom's information requests.
- **Platforms are not prioritising risk assessment processes**, which Ofcom believes are fundamental to proactively identifying and mitigating risks to user safety and which will be a requirement on all services under the OSB, when enacted.

Beyond this, our experience in Year One provided us with broader learnings about how systems and processes-focused regulation work in practice. Most notably:

- **The regulatory model does work**. We witnessed several important regulatory 'successes' in Year One, with companies taking proactive steps to enhance their safety measures on the back of regulatory engagement. The VSP Framework was an instrumental factor behind OnlyFans adopting age verification tools for all new UK subscribers; TikTok establishing an Online Safety Oversight Committee; Vimeo restricting mature and unrated content to account subscribers; and a myriad of other tangible improvements from the VSP sector. These individual instances of success highlight how the VSP Framework can drive improvements in the sector, and they provide an important basis on which providers can continue to improve in the years to come.
- **Transparency is an important regulatory lever**. Many online platforms already publish voluntary transparency reports, with companies choosing how, what, and when they report. Yet these reports provide only a partial account of what is happening inside companies and across the platforms they operate. Thanks to the information-gathering powers of the VSP Framework, we have been able to gather information that goes beyond the voluntary disclosures platforms make and to request further information to help us determine the effectiveness of services' safety efforts¹⁶.
- **Supervisory relationships take time to develop**. For some providers this is their first experience of regulation and of working with Ofcom. Moreover, many providers are unfamiliar with the regulatory model that underpins the VSP Framework – whereby they are empowered to define and execute the measures required to achieve the regulatory outcomes. Regulators have a crucial role to play in helping services understand the Framework and to make progress towards its objectives. This role requires open, constructive, and continued dialogue with services, and as with any relationship that requires trust and understanding, patience is key.

¹⁶ For more information on Ofcom's approach to transparency reporting and its role in online safety, see [here](#).

- **International collaboration can unlock critical synergies:** It is essential to work with other regulators, to both avoid the risk of regulatory arbitrage and to ensure that the rules facing VSPs are consistent. For instance, diverging approaches to the specifications and standards for age verification deployment across countries can create avoidable compliance frictions and duplication. To address these and other concerns, we worked with several counterpart VSP regulators to create the International Working Group on Age Verification, a setting where regulators can cooperate and work towards greater coordination on specifications for the use of age verification measures by the VSPs that we respectively regulate. This international collaboration benefits us as regulators and helps us achieve our respective domestic objectives.
- **Progress is going to be iterative.** VSPs are varied, complex and fast-changing – certainly compared to the relatively well-established sectors that we already regulate. A huge amount of what goes on under VSP providers’ bonnets has never been looked at closely and there remain considerable information asymmetries. Given the novelty of the VSP Framework’s regulatory approach and the fact that we are constantly learning more about harms, service functionalities, and the impact of safety measures, so our regulatory approach must be iterative too. We – both us and the companies that we regulate – are, to some extent, learning on the job about what works and what does not. Regulators should not expect to have all the answers and be able to fix all the problems on Day One.

How we are approaching Year Two

We have sought to capitalise on the learnings from Year One (2022) and to ensure that we continue to drive iterative improvements in VSPs’ safety standards in our strategic priorities for Year Two (2023). Those priorities are to:

- ensure VSPs have sufficient processes in place for setting and revising comprehensive terms and conditions (generally known as Community Guidelines) that cover all relevant harms;
- check that VSPs apply and enforce their Community Guidelines consistently and effectively to make sure harmful content is tackled in practice;
- review the tools VSPs provide to allow users to control their experience and promote greater engagement with these measures; and
- drive forward the implementation of robust age assurance, to protect children from the most harmful online content (including pornography).

In Year Two we continue to be guided by the principles of proportionality, flexibility, and dynamic supervision. We engage intensively with the services that we regulate, and work with them to drive safety improvements. One important difference in our approach in Year Two is how we communicate our findings. This year, instead of producing one all-encompassing end-of-year report, we intend to produce four specific reports over the course of the year, each dedicated to one of our strategic priorities.

In August 2023, we published [the first of these Year Two reports](#), which focused on what we learned about VSPs’ terms and conditions. Through our dynamic supervision and information requests, we have learned that:

- **Users need advanced reading skills to understand VSPs’ terms and conditions.** This means they are not suitable for many users, including children.
- **VSPs’ terms and conditions do cover most types of material harmful to children** but several aren’t clear about when they make exceptions to their rules.

- **Users are unlikely to understand the consequences of breaking VSPs’ rules.** Potential penalties for breaching rules should be made clear to all users in the terms and conditions and this information should be easy to find.
- **Moderators do not always have sufficient guidance on how to enforce VSPs’ terms and conditions.** The quality of VSPs’ internal resources and training for moderators varies significantly. We encourage VSP providers to ensure these resources are clear to help moderators remove harmful content and escalate very serious cases quickly.
- **Some VSP providers have innovative approaches to updating and testing their guidance for moderators and terms and conditions,** but others could do more to make sure their processes are proactive and forward-looking.

[The report](#) also sets out a list of good practices that we observed amongst notified VSPs. These examples might help VSP providers improve their terms and conditions and assist with their implementation of the Framework duties. As noted, this is the first of four reports on our Year Two strategic priorities that we will publish in the coming months. We expect these reports to shed further light on the progress VSP providers are making to enhance user safety, and the good practices that can help them improve further.

In addition to our public reporting, we intend in Year Two to continue external engagement with the policy community and public at large. This engagement – which we have undertaken since the Framework came into being – aims to build trust and understanding in the Framework and in Ofcom, and to promote good practice. As part of it we regularly participate in domestic and international conferences that bring together various VSP Framework stakeholders.

Ultimately, on the basis of our Year One and Year Two work, we have already gathered important insights about how the VSP Framework operates in practice as well as important learnings about this form of regulation writ large. We hope these learnings might also prove useful to Coimisiún na Meán when supervising the implementation of its upcoming code.

The importance of international collaboration

Many of the services that we regulate – and the online safety challenges that we are trying to address – are global in nature. This has implications for how we approach oversight of services under the VSP Framework as well as how we prepare for the broader transition to a new era of online safety regulation.

Co-operating on cross-border VSP issues

As noted, Ofcom regulates VSPs that notify themselves as meeting the statutory criteria for UK establishment. While this oversight role helps us to drive change and safety advancements in the sector, it also means that a range of VSPs that are used by UK-based adults and children are beyond our regulatory remit, and fall under the remit of counterpart regulators in other European countries. As such, to protect individuals in the UK, we depend on regulators like Coimisiún na Meán, just as they rely on us to protect individuals in *their* countries from safety challenges that arise on the VSPs that are notified to us.

In this context it is essential for counterpart VSP regulators to build and maintain close working relationships. By working towards international regulatory consensus on what we as regulators consider to be the appropriate use of safety measures as well as collaborate to create a common understanding of the criteria and standards by which we evaluate the effectiveness of these measures, we can improve safety outcomes and reduce compliance burdens for providers across our jurisdictions. Moreover, by sharing information, where consistent with our statutory rules on information sharing, and working towards coordination of investigations and enforcement, we can

reduce the risk of safety gaps and regulatory arbitrage in the sector. The example of age verification and the corresponding International Working Group outlined in Section 4 is a case-in-point of the need for, and benefit of, international cooperation amongst VSP regulators.

Building a shared vision for online safety regulation

Beyond the immediate challenges and opportunities of supervising the VSP Framework, we recognise that we are transitioning to a new era of online safety regulation, where we can bring to bear a range of innovative policy levers to drive change as well as robust enforcement tools to ensure compliance.

We are conscious that Ofcom is not the only regulator in the world that is on this journey, and we know that we do not have all the answers to the ‘challenging’ questions of online safety regulation. In that context, we see cooperation among international online safety regulators and the multistakeholder policy community as a key ingredient for the success of this new era of regulation. The keystone features of our emergent regulatory frameworks like the OSB and the EU Digital Services Act are novel, and there does not yet exist an international consensus on how these new concepts such as risk management, codes of practice, and mandatory transparency reporting should be implemented. As regulators trying to answer these challenging questions, we have so much to learn from each other, and together we can work towards common understandings of the norms, principles, and standards that will determine how these regulatory tools evolve in the years to come.

For Ofcom, our regulatory philosophy is to aim for international alignment with regulatory partners where appropriate and possible. As in the specific example of VSP regulation, we see many benefits to this approach – for us, our counterpart regulators, the companies we regulate, and ultimately the users that the online safety regulation seeks to protect.

We welcome the fact that Coimisiún na Meán shares a similar commitment to international regulatory cooperation. Through fora like the Global Online Safety Regulators Network; the International Working Group on Age Verification and the various multilateral processes and multistakeholder processes to which we are party, we can together shape a proportionate, effective, and rights-respecting global approach to online safety regulation.

Conclusion

In this submission we have sought to explain our approach to implementing and supervising the UK’s VSP framework over the last two years. Our hope is that our experiences and approach might provide some useful insights for Coimisiún na Meán as it continues its own regulatory preparations.

In the annex we list the various research publications, calls for evidence, and guidance documents that we have issued in recent years that we hope can provide some answers and insights for the specific questions being consulted on. And as always, we will happily provide further details on any aspect of our regulatory approach or experience.

Annex

Ofcom guidance and reports

- **Video-sharing platform guidance: guidance for platforms on measures to protect users from harmful material**
https://www.ofcom.org.uk/_data/assets/pdf_file/0015/226302/vsp-harms-guidance.pdf
- **Ofcom's first year of video-sharing platform regulation: what we found**
https://www.ofcom.org.uk/_data/assets/pdf_file/0032/245579/2022-vsp-report.pdf
- **Regulating Video-sharing platforms | Our first 2023 report: what we've learned about VSPs' user policies**
https://www.ofcom.org.uk/_data/assets/pdf_file/0025/266173/VSP-user-policies-report.pdf
- **Video-sharing platforms: who needs to notify to Ofcom**
https://www.ofcom.org.uk/_data/assets/pdf_file/0023/215456/guidance-video-sharing-platforms-who-needs-to-notify.pdf
- **Video-sharing platforms: Ofcom's plan and approach (letter from Group Director Kevin Bakhurst to notified services)**
https://www.ofcom.org.uk/_data/assets/pdf_file/0024/234177/letter-vsp-plan-and-approach.pdf
- **Repeal of the VSP regime: what you need to know**
<https://www.ofcom.org.uk/online-safety/information-for-industry/vsp-regulation/repeal-of-the-vsp-regime>

Call for Evidence responses

- **Ofcom call for evidence: first phase of online safety regulation**
<https://www.ofcom.org.uk/consultations-and-statements/category-1/online-safety-regulation-first-phase>

Ofcom research

- **The VSP Landscape: Understanding the video-sharing platform industry in the UK (ofcom.org.uk)**
https://www.ofcom.org.uk/_data/assets/pdf_file/0030/245577/2022-vsp-landscape.pdf
- **VSP Tracker Waves 1 and 2: Chart pack (ofcom.org.uk)**
https://www.ofcom.org.uk/_data/assets/pdf_file/0028/245575/2021-22-vsp-tracker.pdf
- **VSP Parental Guidance Research: Summary Report (ofcom.org.uk)**
https://www.ofcom.org.uk/_data/assets/pdf_file/0031/245578/2022-vsp-parental-guidance-research.pdf

- **Understanding how to keep children safe online – Ofcom**
<https://www.ofcom.org.uk/research-and-data/online-research/keeping-children-safe-online>
- **Behavioural insights for online safety: understanding the impact of video sharing platform (VSP) design on user behaviour – Ofcom**
<https://www.ofcom.org.uk/research-and-data/economics-discussion-papers/understanding-the-impact-of-vsp-design-on-user-behaviour>
- **Adult Users’ Attitudes to Age Verification on Adult Sites (ofcom.org.uk)**
https://www.ofcom.org.uk/_data/assets/pdf_file/0029/245576/2022-adult-attitudes-to-age-verification-adult-sites.pdf
- **Families’ attitudes towards age assurance: research commissioned by the ICO and Ofcom (jointly-published with the Information Commissioner’s Office)**
https://www.ofcom.org.uk/_data/assets/pdf_file/0026/245195/DRCF-Ofcom-ICO-age-assurance.pdf
- **The Buffalo attack: Implications for online safety – Ofcom**
<https://www.ofcom.org.uk/research-and-data/online-research/the-buffalo-attack-implications-for-online-safety#:~:text=The%20attack%20was%20livestreamed%20online,to%20content%20related%20to%20terrorism.>
- **Our media literacy research – Ofcom**
<https://www.ofcom.org.uk/research-and-data/media-literacy-research/publications>



Response to Call for Inputs: Online Safety by Coimisiún na Meán

Developing Ireland's First Binding Online Safety Code for Video-Sharing Platform Services

4th September 2023

Section 1: Online Harms

Question 1: What do you think our main priorities and objectives should be in the first binding Online Safety Code for VSPS? What are the main online harms you would like to see it address and why?

Question 2: What types of online harms do you think should attract the most stringent risk mitigation measures by VSPS? How could we evaluate the impact of different types of harms e.g. severity, speed at which harm may be caused? Is there a way of classifying harmful content that you consider it would be useful for us to use?

The first priority and objective of the new Online Safety Code must be to ensure an effective, enforceable code which shows real teeth in addressing online harms. The greatest risk to any such code of practice is that it becomes an ineffectual checklist rather than a living document that genuinely influences and reshapes the behaviours of video-sharing platform service providers, and through this mitigates against clearly identified online harms.

In a previous submission to the drafting of the Online Safety and Media Regulation Act 2022, spunout joined with Cybersafe Ireland, the ISPC and the Psychological Society of Ireland to call for clear definitions as to what constitutes "harmful content" in the context of a future online safety code. We therefore welcome that the Act, and consequently this current consultation, has taken efforts to clearly define and clarify a wide range of categories of online harm.

In particular, we welcome that the Act allows the Online Safety Code to specify areas of online harm beyond the more narrow scope required by the Audiovisual Media Services Directive.

Patron: President of Ireland, Michael D Higgins.

Community Creations Company Limited by Guarantee trading as **spunout** and **50808**

Companies Registration Office Number: 384783 | **CHY (Revenue) Number:** 16212 | **Registered Charity Number:** 20057923

Board of Directors: Tara Doyle (Chairperson), Conor Nolan (Treasurer), Suzanne Mulholland, Barry Ryan, Maria McCann, Daniel Waugh, Conor Healy, Dermot O'Sullivan, Ross Boyd, Aisling Maloney

Copyright © Community Creations 2023



References to bullying, harassment, non-consensual sharing of intimate images were all called for in our previous submission, and we would encourage the Online Safety Code to ensure each of the additional categories enumerated in the Act are given the same attention and opportunities for enforcement as those arising directly from the AMSD.

We do note with concern, however, that there is no direct reference made in the Call for Inputs to hate speech or incitement directed on the basis of gender identity in general and towards trans people in particular. We find this omission concerning in the context of online media regulation, considering the significant and well-documented campaigns of harassment and hostility towards trans service users across a number of social media services in recent years. While sexual orientation is referenced within the definitions of online harms provided, we would strongly urge that the equally important categories of gender identity and expression receive the same level of attention in the final code.

This omission aside, we appreciate that a wide definition of online harms raises the issue of enforcement, and which specific areas should be addressed with the most attention and urgency. There is no easy way to answer this question, as all the identified areas of online harm demand a robust enforcement of an online safety code in order for the code as a whole to be effective. We would, however, strongly encourage that forms of online harm which directly target the wellbeing of an individual, and particularly those which may threaten the life of a person, be given appropriate attention and focus wherever possible.

While our hope and expectation would be that all identified forms of online harm received equal and appropriate priority under the Code, in cases where resources are limited and it is necessary to differentiate, we would encourage the Code to prioritise direct harms such as child sexual exploitation, promotion of suicide or self-harm, incitement to violence against people or groups of people, etc over the regulation of commercial communications, pornography or similar, where the harm caused to individuals is, generally less directly impactful on their life, health or dignity.

This is, of course, not to say that regulation of inappropriate commercial communications or pornographic material should not be a priority for the Code, only that there are areas of online harm with which the impact may be more immediate, and which may therefore require a faster, more effective response.

Patron: President of Ireland, Michael D Higgins.

Community Creations Company Limited by Guarantee trading as **spunout** and **50808**

Companies Registration Office Number: 384783 | **CHY (Revenue) Number:** 16212 | **Registered Charity Number:** 20057923

Board of Directors: Tara Doyle (Chairperson), Conor Nolan (Treasurer), Suzanne Mulholland, Barry Ryan, Maria McCann, Daniel Waugh, Conor Healy, Dermot O'Sullivan, Ross Boyd, Aisling Maloney

Copyright © Community Creations 2023



Section 2: Overall Approach to the Code

Question 4: What approach do you think we should take to the level of detail in the Code? What role could non-binding guidance play in supplementing the Code?

Question 5: What do you think would be the most effective structure for the Code? What are the most important factors we should consider when we decide how to structure the Code?

In weighing up the choice between a more prescriptive, as compared to a high level Code, we refer back to our initial statement that the highest priority must be to ensure a Code that has genuine regulatory teeth. To that end, we believe that a Code which is too high level may only create the illusion of regulation while largely leaving video sharing platform service providers to their own devices. The fact that such a Code is required at all indicates very strongly that a hands-off approach to online safety on the part of the State has not been effective to date.

While there may of course be legitimate need for flexibility or differentiation between service providers of different types, sizes and pre-existing standards of self-regulation, we generally believe that it is the role of an Online Safety Code to provide absolute clarity on the practices and procedures required from all service providers. We do not believe, for instance, that there is a wide range of appropriate response times or enforcement actions which may be taken against online harms. While certain elements of practice may vary from service provider to service provider, the general need for speedy, effective and comprehensive enforcement action against serious online harms must be universal, and clearly understood, across the sector.

Therefore, spunout favour a more detailed Code, if necessary with certain identified exceptions or alternatives as needed on a case-by-case basis, rather than a more general Code which largely continues the existing problems of self-regulation by service providers. For the same reason, spunout approves of the initial decision that there should be a single Online Safety Code, which we believe will better promote clarity and consistency than multiple codes, especially at the beginning of this current regulatory process.

Patron: President of Ireland, Michael D Higgins.

Community Creations Company Limited by Guarantee trading as **spunout** and **50808**

Companies Registration Office Number: 384783 | **CHY (Revenue) Number:** 16212 | **Registered Charity Number:** 20057923

Board of Directors: Tara Doyle (Chairperson), Conor Nolan (Treasurer), Suzanne Mulholland, Barry Ryan, Maria McCann, Daniel Waugh, Conor Healy, Dermot O'Sullivan, Ross Boyd, Aisling Maloney

Copyright © Community Creations 2023



Question 6: How should we design the Code to minimise the potential for conflict and maximise the potential for synergies in how platforms comply with it and the DSA?

We appreciate the potential value in building on the existing requirements of the DSA, especially in such cases where mirrored regulations may increase the likelihood of effective enforcement of both the Online Safety Code and the DSA. However, when designing and implementing the Code, we believe it is imperative that, should any conflict exist between the development of an effective Code and the promotion of synergies with the DSA, that the focus must be on the former. In no case should the Code find itself constrained, or unable to achieve its desired outcomes, due to a perceived need to match too closely with the requirements of the DSA.

While similarities between the two sets of regulations may well make for a simpler regime for service providers to comply with, it is worth noting that the major players in this sector are well-funded and perfectly able to procure the necessary legal and governance advice which may be required to comply with two codes of practice, even in cases where they may not precisely match one another in their requirements. Therefore, ease of compliance with the Code should only be considered a positive where it ensures more effective enforcement outcomes, rather than being a virtue in and of itself - after all, the easiest Code to comply with would likely also be the least effective in actually affecting the behaviour of those it seeks to regulate.

Patron: President of Ireland, Michael D Higgins.

Community Creations Company Limited by Guarantee trading as **spunout** and **50808**

Companies Registration Office Number: 384783 | **CHY (Revenue) Number:** 16212 | **Registered Charity Number:** 20057923

Board of Directors: Tara Doyle (Chairperson), Conor Nolan (Treasurer), Suzanne Mulholland, Barry Ryan, Maria McCann, Daniel Waugh, Conor Healy, Dermot O'Sullivan, Ross Boyd, Aisling Maloney

Copyright © Community Creations 2023



Section 3: Measures to be taken by Video-Sharing Platforms

Question 8: How should we ask VSPS providers to introduce a feature that allows users to declare when videos contain advertising or other type of commercial communications? Should the Code include specific requirements about the form in which the declaration should take? What current examples are there that you regard as best practice?

We welcome the clear intention that the Code will require clear identification of commercial communications on video-sharing platforms. We do not believe that this process needs to be overly complicated, and indeed the measures suggested within the Call for Submissions point in a welcome direction, as does the willingness to act in accordance with CCPC's recent research on influencer marketing.

In general, informing service users that they are viewing a video which contains commercial communications should be as straightforward as possible. In order to ensure that people are informed of the commercial nature of the content they are viewing, the CPCC has recommended the use of a small number of clearly-designated tags on all such content. We agree that the use of straightforward tags such as #advertisement or #paidpromotion would be the best way forward, as compared with vaguer requirements such as the example provided of #workswithX, which does not immediately and clearly identify the content as commercial.

Ideally the number of acceptable tags would be as low as possible to develop consumer familiarity with them as markers, and tags should be displayed ahead of all other tags which may be attached to a video. It should not, for instance, be required that a viewer would have to click into a full list of tags to be made aware that they are watching a paid promotional video.

Patron: President of Ireland, Michael D Higgins.

Community Creations Company Limited by Guarantee trading as **spunout** and **50808**

Companies Registration Office Number: 384783 | **CHY (Revenue) Number:** 16212 | **Registered Charity Number:** 20057923

Board of Directors: Tara Doyle (Chairperson), Conor Nolan (Treasurer), Suzanne Mulholland, Barry Ryan, Maria McCann, Daniel Waugh, Conor Healy, Dermot O'Sullivan, Ross Boyd, Aisling Maloney

Copyright © Community Creations 2023



Question 10: What requirements should the Code include about age verification and age assurance? What sort of content should be shown by default to users who are logged out or in private browsing mode and whose age cannot be verified or assured? What evidence is there about the effectiveness of age estimation techniques? What current practices do you regard as best practice? Where accounts are not age verified should default privacy settings be used, should content default to universal content and should contact by others be more limited?

On the vital issue of age verification and age assurance, spunout had previously cautioned against the Government's setting of the so-called "Digital Age of Consent" (DAOC) at 16. Our concerns were grounded in the counter-intuitive knowledge that a higher DAOC actually makes it more difficult to effectively regulate the content to which children and young people are exposed online. Our argument was based firstly on the observation that the previous DAOC of 13 had not been particularly effective in protecting children aged 12 and under from accessing harmful content online, and that by expanding the illusion of greater protection to those under 16, the State was in fact reducing the practical responsibilities of online service providers in terms of content moderation and child protection.

Rather than requiring providers to put additional work into reducing online harms throughout their platform, our current approach to child protection online shifts the burden of responsibility to parents and young people themselves. While parents and young people themselves should be more engaged in ensuring appropriate online activities, we should never absolve platforms of their responsibility to keep their products safe for children.

Our current approach gives internet service providers an easy way to opt out of creating a safer space for children and young people. In the past, when online platforms have been accused of not doing enough to keep children under the DAOC safe, their response has been to say that children of that age should not be on their platform, at least not without parental consent. And yet in Ireland at the moment the average age children actually first go online is 9 years old. The DAOC has not been effective in preventing under-13s from setting up accounts on popular online platforms where personal information is freely shared.

Generally speaking, when creating a social media or video sharing service account, potential users are asked to provide a date of birth, with parental consent only being a factor where a user self-reports as being under 16. Our current approach arguably risks incentivising children to lie about their age to get online. This seriously undermines our ability to make online and

Patron: President of Ireland, Michael D Higgins.

Community Creations Company Limited by Guarantee trading as **spunout** and **50808**

Companies Registration Office Number: 384783 | **CHY (Revenue) Number:** 16212 | **Registered Charity Number:** 20057923

Board of Directors: Tara Doyle (Chairperson), Conor Nolan (Treasurer), Suzanne Mulholland, Barry Ryan, Maria McCann, Daniel Waugh, Conor Healy, Dermot O'Sullivan, Ross Boyd, Aisling Maloney

Copyright © Community Creations 2023



video-sharing spaces safer for children, as it means it won't be possible to know how many children are actually online.

Regardless of the above, the Digital Age of Consent is currently set at 16 and we accept that this is unlikely to change in the immediate future. However, we would strongly urge the Online Safety Code to draw from the lessons of the DAOC's implementation. Creating ineffective age barriers may appear to 'solve' the problem of young people accessing inappropriate online harms, but the danger is always that we force the problem out of sight and beyond the ability or interest of service providers to effectively solve.

An effective Code would be extremely wary of any pretence that service providers have successfully prevented young people from accessing their services when every piece of evidence indicates an extremely high level of online activity from children and young people below the age of 16. Effective regulation would start from a baseline of assuming young people will be accessing online services of all kinds, and judge success in reducing their access to online harms in line with service provider's demonstrated ability to reduce these harms for all potential users. We are, of course, under no illusions that this is a complex goal; but ensuring safer spaces for everyone is a far more meaningful intervention for young people as compared with ineffectual measures that purport to reduce the number of young people accessing services in the first place.

Patron: President of Ireland, Michael D Higgins.

Community Creations Company Limited by Guarantee trading as **spunout** and **50808**

Companies Registration Office Number: 384783 | **CHY (Revenue) Number:** 16212 | **Registered Charity Number:** 20057923

Board of Directors: Tara Doyle (Chairperson), Conor Nolan (Treasurer), Suzanne Mulholland, Barry Ryan, Maria McCann, Daniel Waugh, Conor Healy, Dermot O'Sullivan, Ross Boyd, Aisling Maloney

Copyright © Community Creations 2023



Question 12: What requirements should the Code have in relation to parental control features? How can we ensure that VSPS providers introduce the mechanism in a user-friendly and transparent way? Can you point to any existing example of best practice in this area? Should parental controls be ‘turned-on’ by default for accounts of minors or where age is not verified?

Question 13: What requirements should the Code contain to ensure that VSPS provide for effective media literacy measures and tools?

We would argue that the issue of parental controls in online spaces cannot be separated from the need for effective media literacy measures and tools. We have stated, in our response to Question 10 above, the risks inherent in transferring responsibility for child safety from service providers to parents. While parents should hold a greater active role in determining the appropriate level of access to online video content by their children, increased responsibility will not be effective without an accompanying increase in online media literacy where needed.

Research from the DCU Anti-Bullying Centre indicates a major digital divide between parents and their children, with as many as 50% of Irish parents reporting that they have insufficient knowledge of online spaces and data protection. Any improvements of parental control features must be cognisant of the major knowledge gaps preventing parents from exercising effective control, and avoid merely creating a false sense of parental security. Clarity, transparency and education around the principles of online safety should be an absolute requirement on service providers when it comes to how they implement systems of parental control. However, such systems should never be seen to reduce the obligations on providers to reduce online harms which young people may face on their platforms, irrespective of parental consent.

Patron: President of Ireland, Michael D Higgins.

Community Creations Company Limited by Guarantee trading as **spunout** and **50808**

Companies Registration Office Number: 384783 | **CHY (Revenue) Number:** 16212 | **Registered Charity Number:** 20057923

Board of Directors: Tara Doyle (Chairperson), Conor Nolan (Treasurer), Suzanne Mulholland, Barry Ryan, Maria McCann, Daniel Waugh, Conor Healy, Dermot O’Sullivan, Ross Boyd, Aisling Maloney

Copyright © Community Creations 2023



Question 14: How should we ask VSPS providers to address online harms in their terms and conditions in the Code, including the harms addressed under Article 28b? How should key aspects of terms and conditions be brought to users' attention? What examples are there of best practice in relation to terms and conditions including content moderation policies and guidelines?

While VSPS providers may choose to implement the requirements of the Code through their terms of services, and to place differing levels of emphasis on certain requirements as per their unique business models, we strongly believe that a minimum set of requirements for restricting online harms be mandated under the terms of the Code.

We welcome the sample bullet points enclosed in the Call for Inputs, setting out prohibitions on criminal or inciting content, clear identification of commercial content, prohibition of harmful commercial content, requirement to age-rate potentially harmful content, and sufficient sanction for users who break the rules.

On the final point, as has been seen in recent months on the service formerly known as Twitter, failure to effectively, consistently and permanently remove users who have broken the terms and conditions of the service user can create a harmful online environment where progress towards reducing online harms goes rapidly into reverse. Therefore, a prohibition against arbitrary restoration of banned service users to their former accounts must form part of an effective requirement to remove users spreading online harm.

Patron: President of Ireland, Michael D Higgins.

Community Creations Company Limited by Guarantee trading as **spunout** and **50808**

Companies Registration Office Number: 384783 | **CHY (Revenue) Number:** 16212 | **Registered Charity Number:** 20057923

Board of Directors: Tara Doyle (Chairperson), Conor Nolan (Treasurer), Suzanne Mulholland, Barry Ryan, Maria McCann, Daniel Waugh, Conor Healy, Dermot O'Sullivan, Ross Boyd, Aisling Maloney

Copyright © Community Creations 2023



Question 15: How should we ask VSPS providers to address content moderation in the Code? Are there any current practices which you consider to be best practice? How should we address automated content detection and moderation in the Code?

On the issue of effective content moderation, we naturally accept that there will always be difficult ‘edge cases’ in which the decision to remove or reinstate content may pose challenges. However, we would caution against overly focusing on these edge cases, compared to the much more important consideration of whether a clear majority of obviously harmful content is promptly removed once identified. While automated systems have a clear role to play in this regard, we would stress the importance of human oversight on these systems to ensure a satisfactory outcome in as many cases as possible.

In this context, we would encourage the Code to view automated content detection as a potentially useful tool for VSPS providers, but not to regard its adoption as an end in itself. Whether or not a VSPS has a functional automated content detection system established within its service is only relevant in as far as this supports a high rate of prompt removal of harmful content. Therefore, VSPS providers should be required to prove that, where they utilise automatic content detection and moderation systems, that they possess clear and effective human oversight procedures needed to ensure a satisfactory level of harmful content removal and provide recourse to users where an automated moderation decision has been made incorrectly.

Furthermore, in order to clearly establish the regulatory teeth of the Code, service providers should be required to prioritise issues raised directly by regulators. The nature of online services of all kinds is that a great number of content moderation decisions will likely be pending at all times. A heavy workload cannot, therefore, be allowed to become an excuse to not fully and promptly engage with issues which have come directly to the attention of a regulator. Clearly establishing the primacy and importance of regulator issues would greatly enhance the salience of thorough Code adoption among VSPS providers, and would be far more effective than simply allowing regulator requests to be viewed as one issue to tackle amongst many.

We would also state that we believe the Code should address the need for a workable minimum timeframe to bind all VSPS providers in terms of responding to issues once raised. At present, response times can vary greatly across the sector. We feel that the Code should aim to implement minimum response times, at the very least for issues of threat to life, risk to children and serious online harms such as intimate image sexual abuse.

Patron: President of Ireland, Michael D Higgins.

Community Creations Company Limited by Guarantee trading as **spunout** and **50808**

Companies Registration Office Number: 384783 | **CHY (Revenue) Number:** 16212 | **Registered Charity Number:** 20057923

Board of Directors: Tara Doyle (Chairperson), Conor Nolan (Treasurer), Suzanne Mulholland, Barry Ryan, Maria McCann, Daniel Waugh, Conor Healy, Dermot O’Sullivan, Ross Boyd, Aisling Maloney

Copyright © Community Creations 2023



Patron: President of Ireland, Michael D Higgins.

Community Creations Company Limited by Guarantee trading as **spunout** and **50808**

Companies Registration Office Number: 384783 | **CHY (Revenue) Number:** 16212 | **Registered Charity Number:** 20057923

Board of Directors: Tara Doyle (Chairperson), Conor Nolan (Treasurer), Suzanne Mulholland, Barry Ryan, Maria McCann, Daniel Waugh, Conor Healy, Dermot O'Sullivan, Ross Boyd, Aisling Maloney

Copyright © Community Creations 2023



Question 16: What requirements should the Code include about procedures for complaint-handling and resolution, including out-of-court redress or alternative-dispute resolution processes? To what extent should these requirements align with similar requirements in the DSA? What current practices could be regarded as best practice? How frequently should VSPS providers be obliged to report to the Commission on their complaint handling systems and what should those reports contain? Should there be a maximum time-period for VSPS providers to handle user complaints and if so, what should that period be?

The timely, effective and satisfactory processing of user complaints by VSPS providers must be a core element of the Online Safety Code. It is only through service user complaints that the true effectiveness of VSPS compliance with the terms of the Code can be demonstrated, and therefore the Code should be strict in setting the minimum standards of complaints handling by service providers. It must be borne in mind, however, that complaints may arise based on a variety of factors, including many which will not be related to the terms of the Online Safety Code. Given the importance of the Code for the overall functioning of a safe online environment, we would suggest that VSPS providers receive a clear obligation to satisfactorily process complaints relating to the terms of the Online Safety Code ahead of complaints unrelated to the issues covered within it.

While we note the statement that this consultation is not seeking views on whether Coimisiún na Meán should accept individual complaints, and welcome that this will be the source of a future consultation, we do wish to once again record our strong belief that the Commission's must adopt an individual complaints mechanism if successful implementation of the Code is to be assured. We note that the European Convention on Human Rights articulates clear rights to fair procedure (Art. 6) and effective remedy (Art. 14) which cannot be said to exist in any regulatory process without a clear and functional mechanism for appeal to the regulator. Without an individual complaints mechanism, the central benefit of this Code (moving away from an era of self-regulation by online service providers) would be significantly undermined.

Patron: President of Ireland, Michael D Higgins.

Community Creations Company Limited by Guarantee trading as **spunout** and **50808**

Companies Registration Office Number: 384783 | **CHY (Revenue) Number:** 16212 | **Registered Charity Number:** 20057923

Board of Directors: Tara Doyle (Chairperson), Conor Nolan (Treasurer), Suzanne Mulholland, Barry Ryan, Maria McCann, Daniel Waugh, Conor Healy, Dermot O'Sullivan, Ross Boyd, Aisling Maloney

Copyright © Community Creations 2023



Question 22: What compliance monitoring and reporting arrangements should we include in the Code?

We strongly recommend that compliance with Ireland’s first binding Online Safety Code must be a matter of high importance for the Board and senior management of any VSPS provider operating in the country. Therefore, at minimum, we believe that full compliance with the Code should be a matter for annual review and sign-off by VSPS Boards of Directors.

This will surely facilitate a greater culture of operating within the Code as compared to a regime entirely made up of ad hoc inspections, whereby enforcement of the Code would be entirely dependent on irregular checks. As with other forms of regulation, a system of regularised compliance reporting bolstered by audit action where necessary is likely to yield better results than either set of actions alone.

Question 23: Should the Code have a transition period or transition periods for specific issues? Which areas touched on in this Call for Inputs may VSPS providers require time to transition the most? What time frame would be reasonable for a transition period?

On the issue of a transition period, we would question the need for an overly long period of adjustment for VSPS given the significant resources of many of the affected organisations, and the extent to which many of the requirements of the Code are likely to match with the provider’s existing stated practices in many areas of content moderation. While some period of transition will of course be required, we would discourage the Code from permitting an overly-long lead-in time which might reduce the momentum of achieving full compliance within a clearly-defined timeframe.

If truly necessary, the Code might consider a first-year ‘grace period’ whereby VSPS providers may state their reasons for non-compliance in certain areas as part of their first annual compliance review. This would have the benefit of clearly identifying areas in which adoption of compliance has not been immediate, giving greater insight into the practical challenges of full compliance as early as possible. However, reporting of compliance and non-compliance with the Code, even with such a grace period, should begin as soon as possible after the Code’s publication.

Patron: President of Ireland, Michael D Higgins.

Community Creations Company Limited by Guarantee trading as **spunout** and **50808**

Companies Registration Office Number: 384783 | **CHY (Revenue) Number:** 16212 | **Registered Charity Number:** 20057923

Board of Directors: Tara Doyle (Chairperson), Conor Nolan (Treasurer), Suzanne Mulholland, Barry Ryan, Maria McCann, Daniel Waugh, Conor Healy, Dermot O’Sullivan, Ross Boyd, Aisling Maloney

Copyright © Community Creations 2023



**Response by the
Advertising Standards Authority for Ireland (ASAI)
to the Coimisiún Na Meán
Call for Inputs: Online Safety
Developing Ireland's First Binding Online Safety Code
for Video-Sharing Platform Services**



ASAI Context: purpose, functions, expertise and deliverables

ASAI – the recognised Irish regulator, highly embedded in the advertising ecosystem

The ASAI is the recognised regulator for advertising in Ireland. Now established for over 40 years, and as a not-for-profit and self-regulatory organisation (SRO), it is independent of any State support or burden on the State/taxpayer through its industry-led funding.

The ASAI, with immense expertise and as a contemporary regulator, is currently enforcing the 7th iteration of its extensive Code. The Code remit incorporates all media including linear and non-linear broadcast, digital (web, social, mobile, in-game ads, influencer marketing (user-generated commercial content), vlogs and blogs, etc.), print, outdoor, cinema, brochures/leaflets, etc.¹

The Code scope and application is broad (covering misleading advertising, taste and decency issues etc.) and comprehensive (covering 14 specific areas including Children, Food (including HFSS rules), Alcoholic drinks, Gambling, Health/Beauty). Code editions are future proofed to encompass industry and societal change. Primary responsibility for compliance with the Code rests with advertisers which means that the ASAI can take action regardless of which medium or platform an advertisement appears on. Media in Ireland, both off and online, support the ASAI Code and its implementation.

The protection of minors remains a category of crucial importance to the ASAI and its Code (7th Edition containing specific reference to children in 72 sections) having introduced initial protections since 1981 at the organisation's foundation. Further details are contained in sections below.

The Code fully recognises this category in society and the added protections that the ASAI Code considers essential for children. The Code in this regard covers a number of broad areas with associated detailed provisions specified in the Code including:

- Protections against matters related to physical, mental or moral harm and/or likely to frighten or disturb children.
- Not to exploit the loyalty, credulity, vulnerability or lack of experience of children
- Promotions to children and related promotional marketing practices
- Marketing communications prohibited from being directed at children or in any way to encourage them to start drinking.
- Marketing communications for food and beverages addressed to children.
- Protections over marketing communications related to gambling.

¹ Appendix I sets out the full remit of the ASAI



The ASAI's mission is to ensure the highest standards of advertising in Ireland, across all media (offline and online) through the enforcement of its Code – in the interests of consumers, business, society and advertising generally - resulting in consistency in marketing communications across all media. The ASAI works with the European Advertising Standards Alliance (EASA) to develop and maintain high advertising standards across Europe.

ASAI – Continuously setting and improving standards

The ASAI endeavours to reflect the needs and sentiments of an ever-evolving society through setting advertising standards, underpinned by the principles of being legal, decent, honest and truthful (based on the International Chamber of Commerce Advertising and Marketing Communications Code²). Essentially, ASAI promotes trustworthy and responsible advertising.

Self-Regulatory Organisation (SRO) Model and Co-Regulation

The SRO self-financing model operated by the ASAI is complementary to legislative controls. It operates flexibly, is more readily adaptable than statute and is easily accessible to all stakeholders/service users. It is more appropriate than legislation for subjective/judgemental areas (decency, discrimination, fear, etc.).

The system of advertising self-regulation has been well established in Ireland, and indeed Europe, for many decades and has been shown to be flexible in its approach to new developments in the advertising eco-system.

The ASAI has a long history of working in the co-regulation space with the Department of Health (Alcohol Marketing Communications Monitoring Body –AMCMB Code).

International relationships and cooperation

The European Advertising Standards Alliance (EASA), co-founded by the ASAI, is the single voice on advertising self-regulation issues in Europe and represents 27 national advertising self-regulatory organisations³. The ASAI's CEO is currently a Vice-Chair of EASA.

EASA and its members develop best practice for advertising self-regulation. It is currently working on developing data driven monitoring projects, already developed and in use by the Dutch, French and UK advertising SROs, and which the ASAI will participate in. This intervention has the potential and capacity to revolutionise monitoring high volume advertising for early detection of advertising breaches and in instances before large scale consumers are exposed to it.

² <https://iccwbo.org/publication/icc-advertising-and-marketing-communications-code/>

³ <https://www.easa-alliance.org/>



Complaints handling access for EU citizens: EASA has operated a cross-border complaints system since 1992 which allows citizens in one member state submit a complaint to their local advertising SRO about advertising on media in another country. The local advertising SRO then ensure that complaint is submitted to the appropriate national advertising SRO who investigates and resolves the complaint under the national advertising Code of the country of origin of the media.

ASAI – 6 Pillars and Governance

The ASAI focuses on 6 primary organisational pillars |

Policy	Code	Awareness
Complaints (41,600)*	Copy Advice (3,150 requests)*	Monitoring (30,000 ads)*

*over a period of our 40+ years' service

In line with the developing growth of online advertising exceeding 40% of all advertising media space, the ASAI has experienced a similar percentage shift in complaints from being related to advertising on traditional media to the digital media including the online platforms.

The ASAI (a not-for-profit Company limited by guarantee) is governed by a Board of 10 non-executive directors, representing the advertising industry (advertising agencies, media and advertisers).

The ASAI provides Alternative Dispute Resolution (ADR) through an independent Complaints Committee of 13 non-industry/industry members and an independent chairperson who adjudicate on complaints (over 8,700* formal and published adjudications and over 6,450* upheld complaints). The ASAI's Review Panel assess requests for reviews of Complaints Committee adjudications.

The ASAI is, as the advertising regulator for all media in Ireland, the organisation that Irish society approach with their complaints about advertising content, including that published by the Irish broadcasters. Our complaints process is complainant-centric, and complainants are not required under the ASAI procedure to have submitted their complaint to either the advertiser or the media before availing of the ASAI service, which is free of charge.

Powerful Impact of the ASAI influence and sanctions

The ASAI has over a 98% success rate in having advertising amended/withdrawn that may be in breach of the Code. The media, including digital media providers, will decline to publish material in breach of the Code.

Adjudications, published by the media, constitute a strong 'name and shame' sanction, with associated brand reputation issues and direct/indirect cost through loss of advertising production costs. The ASAI's formal adjudications, generally published bi-monthly, attract strong media interest; the ASAI



are regularly invited on radio, such as RTE's News at One and Drivetime shows as well as Newstalk and Today FM, and a range of local and regional radio stations, to discuss its adjudications.

The ASAI can impose a compulsory copy advice sanction.

ASAI responses to Call for Inputs Questions

Question 1: What do you think our main priorities and objectives should be in the first binding Online Safety Code for VSPS? What are the main online harms¹³ you would like to see it address and why?

ASAI Comment: In relation to audiovisual commercial communications, and the main priorities and objectives for the first Online Safety Code for VSPS, recognising the cross border nature of the platforms that the code will apply to, the first code for VSPS should reflect the text of the AVMSD and only incorporate additional areas where specifically provided for in the OSMR Act.

Question 4: What approach do you think we should take to the level of detail in the Code? What role could non-binding guidance play in supplementing the Code?

ASAI Comment: As there are different types of content on VSPS (such as user-generated content, paid for commercial communications, commercial communications within user-generated content) a flexible approach is recommended. Providing for the concept of detailed guidance ensures that the Commission can react swiftly to developments requiring regulatory oversight, by the introduction of new guidance as necessary. It is suggested that the Code would require VSPS to provide explanations in the event that the guidance is not followed.

Noting the support in the AVMSD for self-regulation, and the provision in the OSMR ASAI is of the view that the Code should refer to advertising self-regulation and encourage engagement by the VSPS with self-regulatory systems that fulfil the requirements of the AVMSD (to note that the ASAI is firmly of the view that it and the advertising self-regulatory systems in membership of EASA comply with the provisions of Article 4a).

Question 5: What do you think would be the most effective structure for the Code? What are the most important factors we should consider when we decide how to structure the Code?

ASAI Comment: Overly complex structures should be avoided.

Question 7: To what extent, if at all, should the Code require VSPS providers to take measures to address content connected to video content?

ASAI Comment: In relation to commercial communications, and to the extent that this could happen, ASAI would suggest that the Code would require VSPS to take measures to address content connected to video content, but where the video content itself is benign that it would not be removed.



Question 8: How should we ask VSPS providers to introduce a feature that allows users to declare when videos contain advertising or other type of commercial communications? Should the Code include specific requirements about the form in which the declaration should take? What current examples are there that you regard as best practice?

ASAI Comment: Influencer marketing is a continually evolving eco system (as indeed is the advertising ecosystem). Our view is that high level principles are likely to have the most longevity. These can be supplemented with guidance notes. The ASAI Code for example, requires that all advertising be designed and presented in such a way that it is clear that it is a marketing communication. (Code Section 3.31). For influencer marketing we have developed guidance and are currently developing joint guidance with the CCPC. This Guidance will require users to include #ad (or similar) in a clear and unambiguous way and/or to use platform provided tools. Guidance provides the opportunity to differentiate between different VSPS; each will have a different architecture and therefore a detailed 'one size fits all' does not seem to be appropriate. While it would be helpful to users if there were platform-based disclosure tools, the code should also recognize that there are other ways to disclose, and options should be provided to users; a specific reference to existing guidance rather than drawing up new detailed rules would be helpful as it leans into the existing work of the CCPC and the ASAI. In addition, we suggest that the code provides that users should not be disincentivised for choosing one method over another.

Question 10: What requirements should the Code include about age verification and age assurance? What sort of content should be shown by default to users who are logged out or in private browsing mode and whose age cannot be verified or assured?

What evidence is there about the effectiveness of age estimation techniques? What current practices do you regard as best practice? Where accounts are not age verified should default privacy settings be used, should content default to universal content and should contact by others be more limited?

ASAI Response: Where age cannot be verified or assured, it would appear to be an appropriate response to protect minors, content should default to universal content.

The UK Advertising Standards Authority's 100 Children Report may be a useful resource for the Commission - [100 Children Report - ASA | CAP](#)

Question 11: What requirements should the Code have in relation to content rating? What do you consider to be current best practice? What experiences have you had using content rating systems on platforms and do you think they have been effective? What steps could we ask VSPS to take to ensure content is rated accurately by users?

ASAI Comment: In relation to commercial communications, and whether a 'one size fits all' content rating system would be appropriate, the ASAI notes that culture and context are important considerations in deciding whether a specific piece of content is appropriate for a child to view or not. While there may be one rating system, it is possible that it could be applied differently across the EU.

Question 12: What requirements should the Code have in relation to parental control features? How can we ensure that VSPS providers introduce the mechanism in a user-friendly and transparent way? Can you point to any existing example of best practice in this area? Should parental controls be ‘turned-on’ by default for accounts of minors or where age is not verified?

ASAI Comment: In relation to whether parental controls be ‘turned on’ by default for accounts of minors or where age is not verified, ASAI would generally support this approach, as it would be a protective measure for minors. There may well be issues around age verification systems, which respect the privacy of individuals, which the ASAI is not placed to comment on.

Question 14: How should we ask VSPS providers to address online harms in their terms and conditions in the Code, including the harms addressed under Article 28b? How should key aspects of terms and conditions be brought to users’ attention? What examples are there of best practice in relation to terms and conditions including content moderation policies and guidelines?

ASAI Comment: There is probably a tension between terms and conditions covering everything that should be captured in great detail and recognising that individuals may not read all the detail. Notwithstanding this, when an individual agrees to terms & conditions, the VSPS can then rely on their agreement. The code should require that an explainer of the key areas are provided to users via on screen at the time, and should not permit users to dismiss the content until a reasonable period of time has elapsed.

Question 15: How should we ask VSPS providers to address content moderation in the Code? Are there any current practices which you consider to be best practice? How should we address automated content detection and moderation in the Code?

ASAI comment: From an advertising regulatory perspective, it is clearly important that the platforms have robust content moderation processes in place, for the different types of content carried on their platform. ASAI considers that the concept of ‘trusted flagger’ would be helpful if it were extended to areas outside of the DSA, and consider that in keeping with the recognition of and encouragement for self-regulation, advertising self-regulatory bodies established in the EU should be actively encouraged to seek to be a trusted flagger. In addition, while the concept of flagging content to the platforms for them to consider it against their terms and conditions as well as the requirements of the Act, ASAI suggests that the Code should require the platforms to cooperate with relevant bodies, including advertising self-regulatory bodies, in the provision of information, including contact information of users whose content is being flagged. In innovative and fast-moving eco-systems, such as online advertising, precedent is developed via adjudication (in Ireland, through the ASAI Independent Complaints Committee). It is vital that regulators, including self-regulatory organisations, can call to account content creators (be they individuals, sole traders, companies) who create commercial content; they should not be able to remain behind a confidentiality wall, solely created by the platforms’ terms and conditions.

Question 16: What requirements should the Code include about procedures for complaint-handling and resolution, including out-of- court redress or alternative-dispute resolution processes? To what extent should these requirements align with similar requirements in the DSA? What current practices could be regarded as best practice? How frequently should VSPS providers be obliged to report to the Commission on their complaint handling systems and what should those reports contain? Should there be a maximum time-period for VSPS providers to handle user complaints and if so, what should that period be?

ASAI Comment: The ASAI would suggest that consideration be given to having high level requirements that a robust complaints handling process be in place, and provide for further guidance on the details of how that would operate. In this way, the Commission and VSPS providers could have the flexibility to amend the process should such flexibility ultimately be required.

The ASAI would also suggest that if maximum time periods are considered, it is recognised that some complaints might relate to areas that should be prioritised, and flexibility around such time frames might be required.

In so far as the subject matter of the complaint might relate to commercial communication, it is suggested that the code includes reference to the complaints handling alternative dispute resolutions process that exist within the ASAI and other advertising self-regulatory bodies. While it is ultimately up to a consumer if they wish to use these processes, they should be made aware of their existence. It is not suggested that these alternative dispute resolution processes be linked to those that are provided for in the DSA, but are provided for as a separate distinct service.

Question 17: What approach do you think the Code should take to ensuring that the safety measures we ask VSPS providers to take are accessible to people with disabilities?

ASAI Comment: While clearly we would agree that there should be measures, ASAI is not expert enough in this area to offer a view.

Question 19: How do you think that cooperation with other regulators and bodies can help us to implement the Code for VSPS?

ASAI Comment: By way of general comment, the ASAI considers that cooperation with other bodies, including regulators, that have the same broad aim, ensures the best outcomes for all stakeholders.

Specifically in relation to commercial communications, the AVMSD provides at Article 4 1, that

“Member States shall encourage the use of co-regulation and the fostering of self-regulation through codes of conduct adopted at national level in the fields coordinated by this Directive to the extent permitted by their legal systems. Those codes shall:

- a) be such that they are broadly accepted by the main stakeholders in the Member States concerned;*
- b) clearly and unambiguously set out their objectives;*



- c) provide for regular, transparent and independent monitoring and evaluation of the achievement of the objectives aimed at; and*
- d) provide for effective enforcement including effective and proportionate sanctions.”*

The ASAI, in existence for over 40 years, fulfils the requirements of Article 4 1.

As the existing advertising regulator with a remit across both digital/online media and non-digital media, ASAI is well placed to partner the Commission in the implementation of the Code as it relates to commercial communications.

Relevant commercial communications from all EU member states that appear on the VSPS will be subject to the Commission’s Code. However, these commercial communications are also subject to the national rules in place in advertising self-regulatory codes. Context and culture matters when it comes to applying advertising codes. The organisations best placed to understand the culture and context are those in each individual Member State.

In order to ensure compliance with the Code that takes account of these cultural imperatives, ASAI suggests that the Code specifically refers to the European Advertising Standards Alliance and its network of advertising self-regulatory organisations.

At a local level, ASAI has extensive knowledge of the advertising eco-system from a regulatory perspective. Our approach has been to ensure that the ASAI code of practice is future proofed so that as new advertising approaches are developed, the Code can apply without further amendment. Owing to our unique position in the advertising regulatory framework⁴ in Ireland, we can engage with the Commission on information sharing, trend spotting, regulatory developments, education and awareness building.

Question 21: Do you have any views on how requirements for commercial content arranged by a VSPS provider itself should be reflected in the Code?

ASAI Comment: As the VSPS operate across borders with advertisers and users based in multiple EU countries, ASAI would suggest that the Code should reflect the wording of Article 9(1) of the AVMSD.

It is noted that the AVMSD provides at Article 4a 1 that:

“Member States shall encourage the use of co-regulation and the fostering of self-regulation through codes of conduct adopted at national level in the fields coordinated by this Directive to the extent permitted by their legal systems. Those codes shall:

- a) be such that they are broadly accepted by the main stakeholders in the Member States concerned;

⁴ Regulatory framework encompasses statutory requirements, co- and self-regulatory approaches



- b) clearly and unambiguously set out their objectives;
- c) provide for regular, transparent and independent monitoring and evaluation of the achievement of the objectives aimed at; and
- d) provide for effective enforcement including effective and proportionate sanctions

The ASAI and the advertising self-regulatory network in Europe have extensive experience in regulating advertising and marketing communications in the online ecosystem. Our codes of advertising standards are reflective of the requirements in Article 9(1) of the AVMSD.

ASAI considers that the systems in place via the advertising self-regulatory network should be leveraged to support the application of the highest standards in advertising. To this end, we consider that the code should explicitly refer to the requirement for cooperation with and support of advertising self-regulatory bodies that operate in compliance with Article 4a 1 of the AVMSD.

It is noted that the AVMSD at article 9 4 provides that

“Member States shall encourage the use of co-regulation and the fostering of self-regulation through codes of conduct as provided for in Article 4a(1) regarding inappropriate audiovisual commercial communications, accompanying or included in children's programmes, for foods and beverages containing nutrients and substances with a nutritional or physiological effect, in particular fat, trans-fatty acids, salt or sodium and sugars, of which excessive intakes in the overall diet are not recommended

Those codes shall aim to effectively reduce the exposure of children to audiovisual commercial communications for such foods and beverages. They shall aim to provide that such audiovisual commercial communications do not emphasise the positive quality of the nutritional aspects of such foods and beverages.”

It is noted that the OSMR Act provides in Section 139K (5) that

“...an online safety code may prohibit or restrict, in accordance with law, the inclusion in programmes or user-generated content of commercial communications relating to foods or beverages considered by the Commission to be the subject of public concern in respect of the general public health interests of children, in particular infant formula, follow-on formula or foods or beverages which contain fat, trans-fatty acids, salts or sugars.”

In relation to commercial communications for the product categories referred to in the preceding paragraphs, ASAI considered that the Code should require that the VSPS engage with systems that comply with Article 4.1(a).

Question 22: What compliance monitoring and reporting arrangements should we include in the Code?

ASAI Comment: Whilst noting that such reporting is an additional compliance burden, ASAI considers that the Code should provide for structured and ad-hoc reporting. While an annual compliance statement would ensure that compliance and governance arrangements that assure it gets attention



at the appropriate senior level, a more frequent, operational reporting structure might be considered as well. This would provide a mechanism for identifying if issues develop between annual statements. It is suggested that the code provide some flexibility for the Commission to escalate matters and require more frequent reporting if needed. In addition, the ability to require ad-hoc reports should also be provided for.

Question 23: Should the Code have a transition period or transition periods for specific issues? Which areas touched on in this Call for Inputs may VSPS providers require time to transition the most? What time frame would be reasonable for a transition period?

ASAI Comment: ASAI would consider that as a general rule, transition periods are appropriate.

Closing comments

This concludes the submission by the ASAI.

The ASAI respectfully request Coimisiún Na Meán to take full account of all raised in its submission. We remain available to the Commission if clarification or further information is required.

ASAI Submission dated 04 September 2023

Appendix I

Code of Standards for Advertising and Marketing Communications in Ireland

Section 2: Scope and Application

The Code applies to (2.2)

- a) Marketing communications in newspapers, magazines and other printed publications, including free distribution newspapers and magazines.
- b) Marketing communications in posters and other promotional media in public places, including moving images and digital screens.
- c) Marketing communications in brochures, leaflets, circulars, mailings, fax transmissions, emails and text transmissions.
- d) Marketing communications broadcast on television or radio or screened in cinemas or with video, DVD or Blu-ray.
- e) Marketing communications carried on any digital and electronic storage materials, media and/or computer systems including, but not limited to, online advertisements in paid-for space (including banner or pop up advertisements and online video advertisements); paid-for search listings; preferential listings on price comparison sites; viral advertisements; in-game advertisements; commercial classified advertisements; advergames that feature in-display advertisements; advertisements transmitted by Bluetooth; advertisements distributed through web widgets and online sales promotions and prize promotions.
- f) Promotional marketing and sales promotions.
- g) Advertorials.
- h) Marketing communications in non-paid-for space online, under the control of the advertiser or their agent, including but not limited to advertisers' own websites, that are directly connected with the supply or transfer of goods, services, facilities, opportunities, prizes and gifts or which consist of direct solicitations for donations.

Your safety and enjoyment when watching online videos

82

Responses

23:43

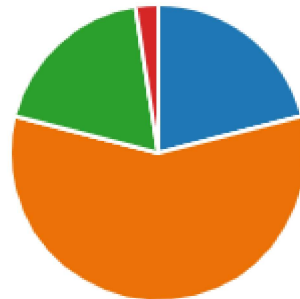
Average time to complete

Closed

Status

1. Knowing how old you are will help us with the survey, can you tell us what age range you fit into?

● 5yrs - 8yrs	17
● 9yrs - 12yrs	47
● 13yrs - 15yrs	15
● 16yrs - 18yrs	2

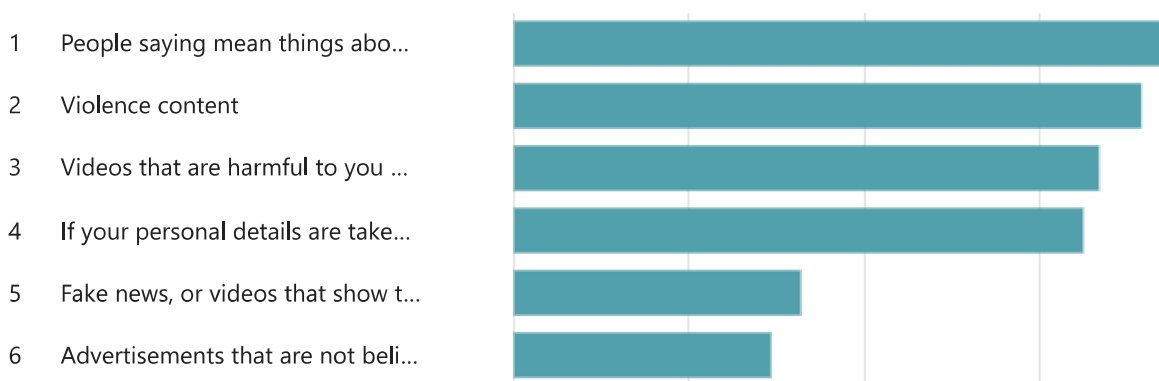


2. When we watch videos online, it is usually for fun, enjoyment and or to find out about something or learn how to do something. But sometimes we can come across videos that aren't nice, they can be frightening, confusing or they can make us feel bad about ourselves.

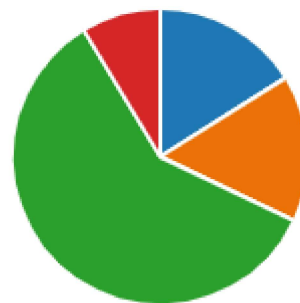
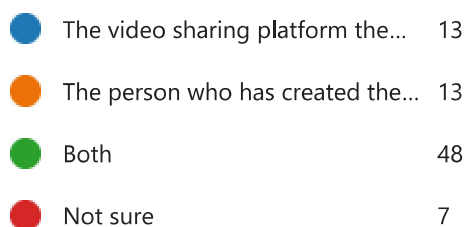
Sometimes it can be hard to know if the person is telling the truth or are they just getting paid to say something?

We would like to know what you think are the worst kind of videos or videos that you think children or young people should not be looking at.

Below you will see different kinds of video content, can you let us know what you think are the worst types by putting them in order? You can drag and drop the different types - putting the worst ones first and the ones that don't bother you as much at the bottom.

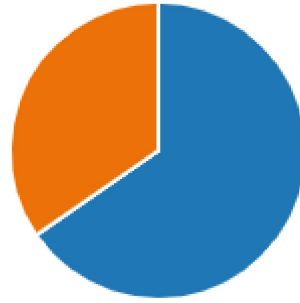


3. Who do you think should be in control or responsible for the types of videos shared online?



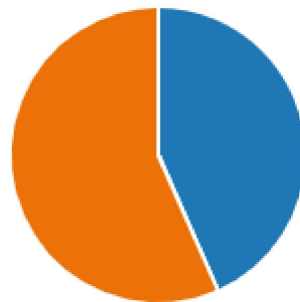
4. Did you know you can report a video if you think it might be bad or unsuitable in some way?

● Yes	53
● No	28



5. If yes, have you ever reported anything?

● Yes	23
● No	30



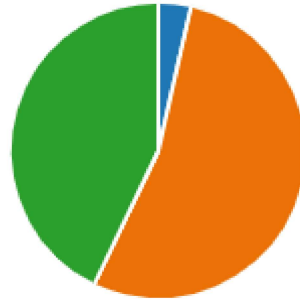
6. Did you find out what happened?

● Yes	2
● No	34
● Other	7



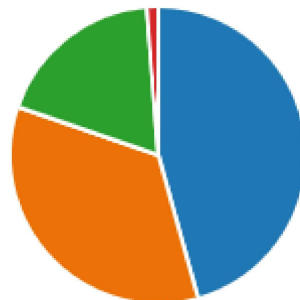
7. If you did find out what happened -were you happy with what happened?

● Yes	1
● No	15
● Other	12



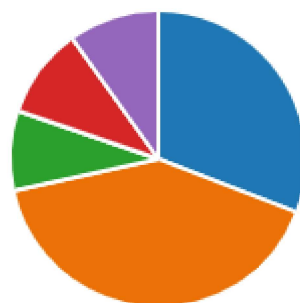
8. Have you ever seen videos that describe what is in the video or who the video is for before you watch it?

● Yes	37
● No	28
● Maybe	15
● Other	1



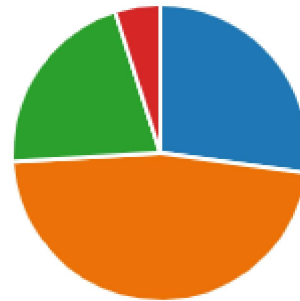
9. If you wanted to know what to expect in a video before you watched it or wanted to know if it was suitable for you to see, what kind of things would help you decide?

● Age-ratings like they use in the ...	25
● A description saying things like: ...	33
● Emoji's, star ratings, thumbs up ...	7
● People's comments on the video	8
● Other	8



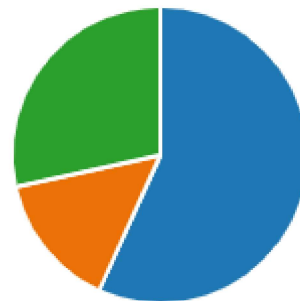
10. Did you know that video sharing platforms like YouTube, TikTok, video games and Instagram already have these descriptions on them?

● Yes	22
● No	38
● Not sure	17
● Other	4



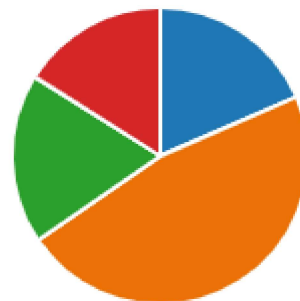
11. If you see or have already seen these kinds of descriptions telling you what might be in the videos before you look at them, do you think it would make you change your mind about watching something?

● Yes	46
● No	12
● Maybe	23



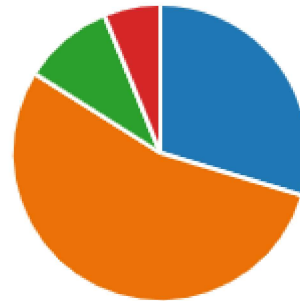
12. Do you think that there is enough information or descriptions about what is in videos before you watch them?

● Yes	15
● No	38
● Not sure	15
● I have never seen any of these d...	13



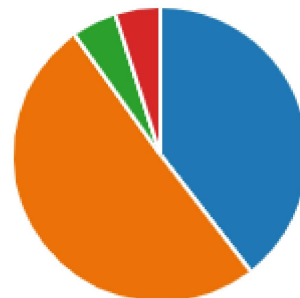
13. Sometimes you can see what people are saying about a video in the comments below them, what do you think about these comments? Are they helpful to you? Please select the option that you think should be the rule.

● Comments should not be allow...	24
● Comments should be allowed b...	44
● I'm not sure	8
● Other	5



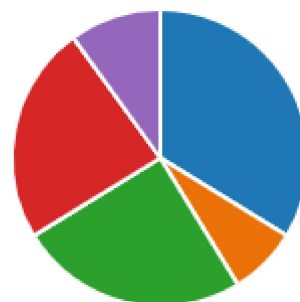
14. Sometimes we see videos of people telling us how good something is so that we might buy it, but what they don't tell you is that they have been paid to say that, and these are really just advertisements that are part of a video that has been made for children on platforms like TikTok, Instagram, and YouTube.

● I think that videos like these sho...	32
● I don't think that videos made f...	41
● I'm not sure	4
● Other	4



15. Can you tell us how you think you should be able to tell the video sharing platform or service how old you are so that you see videos that are most suitable for you?

● By providing proof with a docu...	27
● By letting them guess my age fr...	6
● By just telling them my age and ...	20
● I don't think I should have to tel...	19
● Other	8



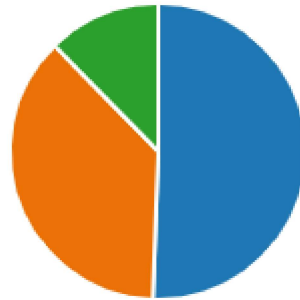
16. Do you know if the videos you are looking at have parental controls?

● Yes	28
● No	23
● I'm not sure	30



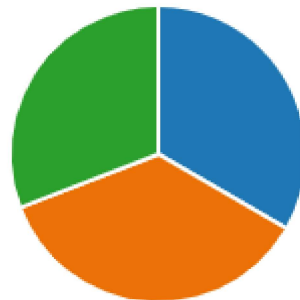
17. Do you think that parental controls should be on all videos that are made for children and young people or should that be up to the parent to put them on?

● Yes, they should be already be t...	41
● No, they should be turned on b...	30
● Not sure	10



18. Did you already know about content feed and how they get the information for that content feed?

● Yes, I already know about conte...	27
● I have heard about content feed...	29
● I did not know anything about c...	25



19. Some children and young people need extra support for reading, writing, hearing difficulties, difficulty seeing or other types of difficulties. If you need extra help for deciding what videos to look at, are there any rules or controls that you would like to see some of the video sharing platforms or companies put in place that might make your online life safer and more enjoyable? Please let us know by adding your comment to the box below.

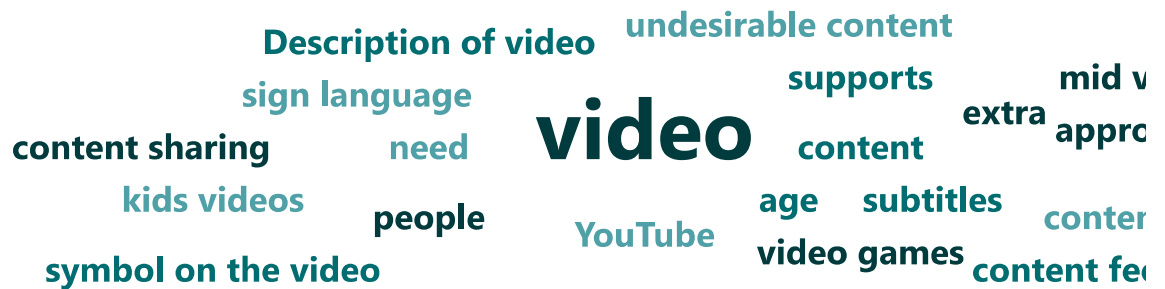
34
Responses

Latest Responses

"I think that there should be automatic subtitles to be turned o...

"Have a section for easy access wirh supports for anyone with e...

11 respondents (32%) answered **video** for this question.



Developing Ireland’s First Binding Online Safety Code for Video-Sharing Platform Services

A ‘call for inputs’ document intended to inform a future consultation by Coimisiún na Meán (the “Commission”) on a draft Online Safety Code

Submission by the partners of the Irish Safer Internet Centre

(Hotline.ie, ISPCC, National Parents Council, Webwise)

4 September 2023



The Irish Safer Internet Centre

This submission is put forward by the Irish Safer Centre. The Irish Safer Internet Centre partner organisations work towards a shared mission of making the internet a safer and more inclusive place for children and young people. A partnership of **Hotline.ie**, **ISPCC**, **National Parents Council Primary** and **Webwise** coordinated by the **Department of Justice** and co-funded by the **European Union**.

The Irish Safer Internet Centre (SIC) has three main pillars:

- Awareness: Webwise
- Helpline: Childline and National Parents Council
- Hotline: Hotline.ie

Additionally, as one of 31 Safer Internet Centres of the InSafe-INHOPE Networks we contribute to the Better Internet for Kids (BIK) core service platform to share resources, services and practices between the European Safer Internet Centres and advice and information about a better internet to the general public. In line with the European Commission's Better Internet for Kids+ Strategy, the key vision behind the BIK core service platform is to create a better internet for children and young people.

Therefore, this submission will focus on the rights and needs of children and young people (minors) in respect of the proposed online safety code for VSPS providers. Our response to this consultation is focused on the relevant operational areas of expertise within the partner organisations of the Irish Safer Internet Centre. Responses to relevant questions are outlined clearly below.

This submission is supported by responses from 11 students from the Webwise Youth Advisory Panel (see attached Appendix 1) to the suggested survey in Appendix 2 and includes responses and views from Webwise Youth Advisory Panel discussions on the Online Safety and Media Regulation Bill during Webwise Youth Advisory Panel meetings from 2020 onwards.

Also included in this submission is responses from a recent survey conducted by the National Parents Council (see attached Appendices 2,3,4). The survey issued in August 2023 includes the collected responses from 595 parents and 82 children and young people aged between 5 and 18 years old.

Question 1: What do you think our main priorities and objectives should be in the first binding Online Safety Code for VSPS? What are the main online harms¹³ you would like to see it address and why?

Main Priorities and Objectives

The Irish Safer Internet Centre believes that the main priorities and objectives of the first online safety code for VSPS providers should ask designated services initially as a principle to proactively prioritise and resource appropriately the safety rights and needs of all its users equally, with enhanced protections for minors, whilst also supporting them to develop into digitally competent citizens who are able to participate meaningfully in the digital environment.

In respect of minors, it is important that it is recognised that children's rights are protected, respected and fulfilled in the digital environment (as they are in the offline world) as outlined in the UN's General Comment No. 25 on children's rights in relation to the digital environment.¹ Children have a right to be protected online and this must be balanced with their right to participate; their right to access information; their right to freedom of expression, etc.

Main Online Harms

The Irish Safer Internet Centre recommends that the harms as per Article 28 (b) of the Audio-Visual Media Services Directive (AVMSD) and the 2009 Act as amended are the harms to be prioritised by this first online safety code whilst also giving consideration to the provisions of the Digital Services Act (DSA) which could enhance and strengthen the code.

According to the most recent, comprehensive online safety research on internet use by children in Ireland conducted on behalf of the National Advisory Council for Online Safety²; 13% of children overall, and one in five (21%) of 15–17-year-olds has experienced something that bothered or upset them in some way, making them feel uncomfortable, scared or that they shouldn't have seen it. People being nasty to each other (24%) and bullying (22%) stand out as the most mentioned things that upset young people.

A quarter of all girls in the survey (26%) listed people being nasty to each other as the issue that most frequently upsets them. This type of content is reflective of what Article 28b (b) of the AVMSD provides for and as per Section 139A, Subsection 3 (a) of the Online Safety and Media Regulation Act 2022. Inappropriate or disturbing videos and photos is the next most significant issue reported by one in five children (19%), followed by cruelty to animals online.

¹<https://www.ohchr.org/en/documents/general-comments-and-recommendations/general-comment-no-25-2021-childrens-rights-relation>

²<https://www.gov.ie/en/press-release/b994b-minister-martin-launches-comprehensive-online-safety-research-on-internet-use-by-children-and-adults-risks-they-face-and-how-they-respond/>

The report also noted exposure to harmful content online. Overall, more than one in four children (26%) report seeing potentially harmful content online in the last year making this the most reported type of online risk that children encounter.

The following forms of harmful content online seen by children and captured in the 2014 Net Children Go Mobile study³ are as relevant today and even more prevalent - "forms of harmful online content seen by children are hate messages (20% compared to 15% in 2014); gory or violent images (18%); experiences of taking drugs (16% vs. 7% in 2014); self-harm sites (13% compared to 9% in 2014) and sites promoting ways to be thin (11%); 9% say they have also seen sites that depict ways of dying by suicide".

According to members of the Webwise Youth Advisory Panel, the following harms should be addressed:

"Not allowing hate content."

"Children - cyber bullying.....young people/teenagers - misinformation, online harassment"

"The spread of harmful content"

"Bullying has a major negative impact on people's mental health, and fake news spreads false information that can be potentially dangerous."

"Being on apps or watching things that are not for their age group and then being brought to seeing content that could make them feel uncomfortable or make them then feel they have to act like the people on these apps or videos act."

Eleven members of the Webwise Youth Panel also completed the suggested survey provided in Appendix 2. The majority indicated they feel quite safe when watching videos online or using apps, many did indicate it was platform dependent and there is potential to encounter inappropriate content.

On the topic of what concerns them online; 6 students indicated they were not concerned or 'not really' concerned. For students that were concerned, here's what they had to say about videos that concern them online:

"Pornography and hate speech is what I have been most exposed to."

"Yes, homophobic, sexist and racist videos."

"Yes. Videos of fights"

In conjunction with these insights from children and young people, the Irish Safer Internet Centre recognises that the heinous crime of child sexual exploitation and abuse online, must continue to be a priority. The proposed EU regulation laying down rules to prevent and combat

³ <https://netchildrengomobile.eu/reports>

child sexual abuse is currently under consideration.⁴ It seeks to provide legal certainty to providers as to their responsibilities to assess and mitigate risks and, where necessary, to detect, report and remove known and new child sexual abuse material as well as child solicitation on their services in a manner consistent with the fundamental rights laid down in the Charter of Fundamental Rights and Freedoms and as general principles of EU law.⁵ The code for VSPS providers would need to consider the intentions of this impending regulation.

Question 2: What types of online harms do you think should attract the most stringent risk mitigation measures by VSPS? How could we evaluate the impact of different types of harms e.g. severity, speed at which harm may be caused? Is there a way of classifying harmful content that you consider it would be useful for us to use?

Online risk to children has been classified according to the 4Cs of content, contact, conduct and contract risks (Livingstone & Stoilova, 2021).⁶ The classification offers the foundations of a better understanding of online risk to children. Policymakers can use it to identify what risks matter and why, what evidence supports them, and how they fit within or fall outside existing regulatory frameworks. This classification was developed in response to emerging and evolving risks for children online being overshadowed by more prevalent ones of cyberbullying, grooming and child sexual exploitation. Providers ought to have a level of familiarity with this classification via engagement with the European Commission's self-regulatory initiative the 'Alliance to better protect minors online'⁷.

The 4Cs: Classifying Online Risk to Children

Recognising that online risks arise when a child:

- engages with and/or is exposed to potentially harmful CONTENT;
- experiences and/or is targeted by potentially harmful CONTACT;
- witnesses, participates in and/or is a victim of potentially harmful CONDUCT;
- is party to and/or exploited by a potentially harmful CONTRACT.

The 4Cs classification also distinguishes between aggressive, sexual and value risks, as this is helpful in retaining a balanced view of the range of risks that children can encounter. The authors note that risks to the values that shape childhood and society are increasingly prominent.

⁴ <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=COM%3A2022%3A209%3AFIN>

⁵ <https://www.coe.int/en/web/children/-/launch-of-an-interdisciplinary-outcomes-report-on-the-potential-implications-of-the-eu-s-proposal-for-a-regulation-to-prevent-and-combat-child-sexual-abuse>

⁶ https://www.ssoar.info/ssoar/bitstream/handle/document/71817/ssoar-2021-livingstone_et_al-The_4Cs_Classifying_Online_Risk.pdf?sequence=4&isAllowed=y&lnkname=ssoar-2021-livingstone_et_al-The_4Cs_Classifying_Online_Risk.pdf

⁷ <https://digital-strategy.ec.europa.eu/en/policies/protect-minors-online>

It was also generally agreed that, to be useful, risk classifications should prioritise:

Flexibility – the classification has to be broad and flexible so that new risks can be added when needed or when we need to refer to different groups of children or address stakeholders.⁸

Clarity – the risks should not overlap with each other and they should map readily onto the reports from children or practitioners about problematic experiences. Recognising that this is a complex domain, the call was also to avoid oversimplification, recognising ‘hybrid threats’ that could be classified in more than one domain (e.g. identity theft could be linked to contact, conduct or contract risks depending on the circumstances; online pressures relating to body image can have both sexual and value dimensions).⁹

Cross-cutting risks: Some risks relate to most or all of the four categories and can have multiple manifestations across the different dimensions (aggressive, sexual, values). These include online risks relating to privacy, physical or mental health, inequalities or discrimination.¹⁰

The authors suggest that the classification system should map onto the actual problems reported by children or encountered by practitioners. They should also resonate with audiences (parents, policymakers, etc.) when risk-related work is made public.

They also note that it is also important to see risk as only one of the dimensions of children’s online experiences, alongside opportunities and among many factors that intersect to influence children’s outcomes.¹¹ Indeed, while the digital environment affords children a range of risks, it also offers many opportunities to benefit, and this merits a parallel analysis. If society becomes overprotective, it can inadvertently undermine the very opportunities for which society provides children with internet access.

During consultations and discussions with the Webwise Youth Advisory Panel in 2020, one teen noted:

⁸ https://www.ssoar.info/ssoar/bitstream/handle/document/71817/ssoar-2021-livingstone_et_al-The_4Cs_Classifying_Online_Risk.pdf?sequence=4&isAllowed=y&lnkname=ssoar-2021-livingstone_et_al-The_4Cs_Classifying_Online_Risk.pdf Page 9

⁹ https://www.ssoar.info/ssoar/bitstream/handle/document/71817/ssoar-2021-livingstone_et_al-The_4Cs_Classifying_Online_Risk.pdf?sequence=4&isAllowed=y&lnkname=ssoar-2021-livingstone_et_al-The_4Cs_Classifying_Online_Risk.pdf Page 9

¹⁰ https://www.ssoar.info/ssoar/bitstream/handle/document/71817/ssoar-2021-livingstone_et_al-The_4Cs_Classifying_Online_Risk.pdf?sequence=4&isAllowed=y&lnkname=ssoar-2021-livingstone_et_al-The_4Cs_Classifying_Online_Risk.pdf Page 11

¹¹ <http://globalkidsonline.net/tools/guides/framework/>

“I think that the [OSMR] bill should hold companies accountable for the spreading of fake news because I genuinely believe that it is one of the most predominant issues currently and the possibility of a post being fake news needs to be outlined to the user.”

The National Parents Council, a partner in the Irish Safer Internet Centre carried out a survey with parents and children and young people to inform this call document.¹² The children and young people who contributed to the survey ranked “people saying mean things about other people and bullying them” as their number one concern, followed closely by violent content. Parents ranked their top three online harms as follows: 72% of parents ranked sexual imagery and abuse; 35% of parents ranked bullying behaviour; 19% ranked the promotion or encouragement of eating disorders, self-harm or suicide” (NPC survey 2023)

The recommendation of the Irish Safer Internet Centre is that the code ought to adopt a principle of proportionality based on the severity of the risk it proposes to mitigate against.

Question 3: Do you have reports, academic studies or other relevant independent research that would support your views? If you do, please share them with us with links to relevant reports, studies or research.

The Irish Safer Internet Centre has referenced other documents in footnotes throughout the body of this call document and has made available additional surveys and insights via accompanying appendices.

- **Research Report: Bystander Behaviour Online Among Young People in Ireland**

Research shows that cyberbullying is a significant issue encountered online by children in Ireland. Researchers and educators recognise the importance of the role of peer bystanders in bullying situations, but more research is needed in this regard within an Irish context. Moreover, there appears to be a general lack of literature on the role of bystanders in cyberbullying situations. Therefore, this research study commissioned by Webwise was conducted by DCU Anti-Bullying Centre and aims to explore online bystander behaviour among young people in Ireland. A sample consisting of 212 students aged 13 to 17 years completed an online survey including questions regarding participants' use of the internet and digital devices and bystander behaviour.

- Cyberbullying is frequently witnessed online, with 45.3% of students surveyed report witnessing some kind of mistreatment online over the last months, being therefore cyberbullying bystanders.
- Various forms of direct verbal abuse are the most common online. From those who witnessed cyberbullying, 64.6% reported name calling, and mockery or insults were also witnessed by 63.5% of the bystanders.

¹² See Appendices 2 and 4

- The space where cyberbullying most often takes place is social media. Of the bystanders, 60.4% reported having witnessed cyberbullying on a social network.
- Among those who reported witnessing cyberbullying, 31.3% said a stranger started it and 25% said other strangers joined in.
- Participants are in general aware of protective mechanisms provided by social networks, and report using those mechanisms to protect themselves, but not so much to help others. The most common mechanism for helping other people is the report button used by 14.2% of the sample.

Full report available here: webwise.ie/saferinternetday2023Report

- Livingstone, S., & Stoilova, M. (2021). ***The 4Cs: Classifying Online Risk to Children. (CO:RE Short Report Series on Key Topics)***. Hamburg: Leibniz-Institut für Medienforschung | Hans-Bredow-Institut (HBI); CO:RE - Children Online: Research and Evidence. <https://doi.org/10.21241/ssoar.71817>
- National Advisory Council for Online Safety: [**Report of a National Survey of Children, their Parents and Adults regarding Online Safety**](#) (2021)
- Smahel, D., Machackova, H., Mascheroni, G., Dedkova, L., Staksrud, E., Ólafsson, K., Livingstone, S., and Hasebrink, U. (2020). ***EU Kids Online 2020: Survey results from 19 countries***. EU Kids Online. Doi: 10.21953/lse.47fdeqj01of0

<https://www.eukidsonline.ch/files/Eu-kids-online-2020-international-report.pdf>¹³

Survey results from 19 countries. This report maps the internet access, online practices, skills, online risks and opportunities for children aged 9–16 in Europe. Teams of the EU Kids Online network collaborated between autumn 2017 and summer 2019 to conduct a major survey of 25,101 children in 19 European countries.

- ***Global Kids Online 2020***

<https://www.unicef-irc.org/publications/1059-global-kids-online-comparative-report.html>

This report gives important insights into Children’s Internet Access; Parental Mediation and Support; Online Activities; Digital Skills; and Children’s Reporting of Online Risks.

- We Protect: [**Child ‘self-generated’ sexual material online: children and young people’s perspectives**](#)
- 5 Rights: [**Pathways: How digital design puts children at risk**](#)

¹³ Smahel, D., Machackova, H., Mascheroni, G., Dedkova, L., Staksrud, E., Ólafsson, K., Livingstone, S., and Hasebrink, U. (2020). EU Kids Online 2020: Survey results from 19 countries. EU Kids Online. Doi: 10.21953/lse.47fdeqj01of0

A submission on behalf of the Irish Safer Internet Centre to inform a future consultation by Coimisiún na Meán on a draft Online Safety Code

Question 4: What approach do you think we should take to the level of detail in the Code? What role could non-binding guidance play in supplementing the Code?

Approach Code should take

Option 3 – a mixed approach

The Irish Safer Internet Centre believes that a mixed approach may be the best option to take initially and perhaps work in a review clause at a particular juncture, perhaps as per Section 46R (3) of the 2009 Act as amended. This would be a similar approach to Australia and the UK and the Data Protection Commission for compliance to GDPR; a high-level code supported with non-binding guidance. The approach the code will take will also need to allow for emerging research to continually inform the code. This is a hugely complex and technical area and such an approach could allow greater flexibility in working through identified gaps and perhaps mitigating against unintended consequences.

Due to the lack of transparency in terms of the volume and type of online harms presenting to VSPS providers; the volume and type of reports being made; the processing and handling of such reports; the lack of insights into the mitigating and aggravating factors into online harms and how they are dealt with; and to accommodate the ever evolving and emerging harms a mixed approach could be prudent in order for all providers to set themselves up for a level success, initially.

Section 139B of the 2009 Act as amended allows for the proposal and consideration of other harmful online content so an approach that is not overly prescriptive could also support this.

Designated services may find it difficult to comply if something is overly prescriptive, leading them to develop bespoke solutions that won't allow flexibility for emerging and evolving amendments and enhancements to the code as it is revised, ultimately proving the code unworkable which would not be in the best interests of anyone.

Where relevant designated services are directed to minors and where relevant designated services are used by minors - although not necessarily directed to them but there is an awareness of minors using them - then the code should seek to use a child rights by design approach.

Child rights impact assessments also ought to be incorporated into the code – this would ask relevant designated services to carry out child rights impact assessments on their service offering including their terms and conditions of use. The Digital Futures Commission published a child rights impact assessment toolkit that would be a useful reference to inform this code.¹⁴ Compliance to the code must be demonstrative in nature supported by tangible examples.

Role of Non-Binding Guidance

¹⁴ <https://digitalfuturescommission.org.uk/wp-content/uploads/2021/03/CRIA-Report.pdf>

The Irish Safer Internet Centre is of the opinion that non-binding guidance could play an important role in supplementing a high-level code outlining 'what' must be done and such guidance offering suggestions of 'how' it could be done. Each VSPS provider will be built on different technologies, some having invested more in online safety than others, non-binding guidance would be able to point to good practice as a means of setting expectations and standards in terms of how to adhere to the code and reduce the need for interpretation to the minimum. Non-binding guidance can support the interpretation of the code with tangible and descriptive examples/scenarios.

Question 5: What do you think would be the most effective structure for the Code? What are the most important factors we should consider when we decide how to structure the Code?

Structure of the Code

The Irish Safer Internet Centre believes structuring the code thematically will allow a VSPS provider to apply the relevant sections that pertain to it in a systematic manner and ought to allow more readily for future harms and/or other provisions to be accommodated within the code and perhaps more workable and avoid the situation of the code being repetitive for each harm. The structure could perhaps suggest different practices and define minimum standards in each thematic area based on the severity of the harms it addresses. There are voluntary codes currently in place whereby various companies have signed up to adhere to.¹⁵¹⁶¹⁷

Important Factors for the Code;

- The background to and purpose of the code and expected outcomes.
- The VSPS providers to whom the code applies to and the rationale for same, including exemptions, if any.
- The end-users protected under the code and who can avail of its provisions, including any exemptions and/or special categories of end-users, if any.
- The code ought to oblige each VSPS provider to indicate the presence of each category of risk on its platform as per the 4Cs: Classifying Online Risk to Children (Livingstone, Sonia; Stoilova, Mariya) as mentioned earlier in this call document; the prevalence of such risks as they learn from proactively and reactively managing these harms; the mitigating factors/policies employed to address these risks (proactively); any aggravating policies or conditions; the actions taken in addressing these risks (reactively).

¹⁵ [In May 2016, Facebook, Twitter, YouTube and Microsoft agreed with the European Commission a Code of Conduct on Countering Illegal Hate Speech Online, Law Reform Commission Report, 2016](#)

¹⁶ [EU Code of Conduct on Countering Illegal Hate Speech Online](#)

¹⁷ <https://digital-strategy.ec.europa.eu/en/policies/code-practice-disinformation>

- The VSPS providers to outline what, if any, education-prevention, detection of predatory behaviour and early intervention initiatives are undertaken to ensure end-users know the existence of these policies; the location of these policies; and accessibility of these policies and for the code to mandate the same. This can also include how VSPS providers educate their end-users to stay safe online.
- Any education initiatives undertaken to enhance the digital literacy of its end-users.
- The complaints handling system that the VSPS provider has in place; the process and indicative timeline to deal with a complaint and what is in place to appeal a decision including the process and indicative timeline for the same.
- That VSPS providers subject to the code demonstrate how they purpose to comply with the provisions of the code.

Question 6: How should we design the Code to minimise the potential for conflict and maximise the potential for synergies in how platforms comply with it and the DSA?

The code needs to be designed with the provisions of the DSA as a key consideration to support maximum adoption and compliance. A key feature of the code will be outlining which providers are subject to which parts of the code as per the distinct provisions of the 2009 Act as amended, the AVMSD and the DSA, where synergies exist and where the delineation is, and why. Noteworthy key feature of the DSA pertinent to this online safety code is that of content moderation and accountability.

In terms of moderating content including notice and takedown, there are opportunities for synergy between the proposed super complaints mechanism in the 2009 Act as amended and the proposed trusted flagger scheme in the DSA. Both are looking to experts to flag content of concern. However, one is for the purpose of review prioritisation and the other is to look at content of concern systemically. The objectives of both could be aligned.

Synergy could also be found in how the code addresses online harms. The code will also have to consider how providers subject to the DSA in this jurisdiction will also have to comply with definitions of online harms in other national legislation, such as provided for in the 2009 Act as amended amongst others, providing there is synergy; where there isn't, then EU law such as the DSA will supersede national law which may pose additional considerations for the code and its provisions and by extension, compliance. It will be important that all harms are clearly defined in the code. Whilst the DSA and the AVMSD are principle-based when providing for online harms, the 2009 Act as amended could offer more substantive definitions to support compliance.

Utilising the provisions within the DSA in respect of recommender systems would support the drive for transparency into how such systems amplify harmful content and allow proliferation of such content on their platforms. Prolonged use of social media and exposure to harmful content has been shown to have a potential negative impact on the end-user's mental health.

Although this area requires more research. Such an approach would be child rights by design where children have a right to exercise agency over their online consumption and it's not a case of recommender systems manipulating them into viewing certain types of harmful content that we know generates better engagement.

The Irish Safer Internet Centre strongly supports the approach that the design of the code would take into account provisions of the DSA where practicable.

Question 7: To what extent, if at all, should the Code require VSPS providers to take measures to address content connected to video content?

When tackling online harms, and especially where the content might not be manifestly illegal or criminal in itself, context can be extremely important. For example, The I-KiZ – Centre for Child Protection on the Internet Paper “Combat of the Grey Areas of Child Sexual Exploitation on the Internet¹⁸”, provides insights on how the context material is placed in can make innocent imagery exploitative. Below an example excerpt from the aforementioned Paper of a real-life case outlining how comments made on an image can sexualise a child (same would apply in principle to videos or any other form of content):

“The depiction shows an under-age boy in swimwear at the beach. This is a typical everyday scene, which does not at first suggest a sexual connotation, but which was however appropriated using service-specific functions and structures. The depiction was posted on the public group of a social network and accessed a thousand times. Users posted many similar pictures.

The following elements sexualised the depiction and made the child into a sexual object:

Title/theme: “NAKED BOYS & GIRLS SEX?”

The title of the group makes it clear users with a sexual interest in children are displaying and using pictures here. This not only harms the right of the child to his image, the child shown is also reduced to a sexual object. Comments: “Hot little fucker”, “Sexy boy yummm”.

Users commented on the depiction using sexualised language and “socially acceptable” sexualised language. In a family context, the comments would have been characterised by respect and would have related to the activities of the child. Here, the

¹⁸ <https://childrens-rights.digital/hintergrund/index.cfm/topic.324/key.1585>

sexual interest is formulated, and the child is reduced to a desired sexual object.

Reinforcing factors:

The activities of commenting users and group members act as reinforcement. Many used depictions of children as profile images, which in this real-life context can indicate a sexual interest in children. Screening of likes, comments, friend lists and memberships of other groups furthermore revealed that many users were networked via several groups and profiles which collected everyday depictions of semi-clad children to which they obviously had no personal relationship.”

If one was to ignore the context the image was in, and the comments made on it then it could not be actioned as the image itself is legal. As such, context is a key component in assessing material. This example was of child sexual exploitation, but the same thing stands to other types of content, such as cyberbullying where it is heavily context dependent.

Additionally, the Irish Safer Internet Centre partner NPC survey, for the purpose of this Call for input, aimed at both parents and children and young people, revealed:

“70% of parents thought that comments should be disabled for videos aimed at children, and 22% felt that the comments should be effectively monitored. The remainder of parents were unsure how they felt about this.

54% of the young people surveyed felt that comments should be allowed but they should be monitored.” (NPC survey 2023)

Question 8: How should we ask VSPS providers to introduce a feature that allows users to declare when videos contain advertising or other types of commercial communications? Should the Code include specific requirements about the form in which the declaration should take? What current examples are there that you regard as best practice?

Sponsored videos have rapidly emerged as an important marketing tool as video-sharing platforms and the popularity of video influencers have grown. The Code should look to ensure that a consistent feature for VSPS providers is introduced across all platforms that places a stringent requirement on users to declare when videos contain advertising and/or commercial communications. It should include a specific requirement for what form the declaration should take. This should be clear, concise, transparent and easy for children and young people to understand.

Parents were asked (NPC survey 2023) if they thought sponsored content should be clearly labelled and regulated to ensure that children can distinguish between regular content and advertisements, OR if they believed that sponsored content should not feature at all in videos

aimed at children and such content should be completely separate from videos meant for young audiences.

“85% of parents believed that sponsored content had no place in videos aimed at children.

One parent commented:

“There should be no advertising whatsoever to minors online, not only things deemed generally inappropriate but also harmful to the individual or unhealthy, which varies widely from person to person. There is no way to fully monitor the damage so it should not be considered at all, it should all be banned for children.”

“39% of the young people surveyed thought that it should be very clear and obvious to them when products or services were being promoted, but 50% felt that these promotions had no place in video content aimed at children or younger people.

One young person commented:

“They should say if their video is just really an ad to get me to buy something”

Question 9: How should we ask VSPS providers to introduce and design a flagging mechanism in the Code? How can we ensure that VSPS providers introduce the mechanism in a user-friendly and transparent way? How should we ask VSP Providers to report the decisions they’ve made on content after it has been flagged? To what extent should we align the Code with similar provisions on flagging in the DSA?

Key defining features of a flagging mechanism should be outlined, on the principle of minimum standardisation of fundamental aspects that are universally applicable to all VSPS. This can be drawn for example from best practice examples such as Youtube’s priority flagger program.

By establishing basic expectations to ensure a minimum standard and thus a level of uniformity across VSPS, and providing further guidance such as best practice example. Have on-platform accessibility by design (e.g. include voice-activated option) reporting tools enabling complaints, whilst establishing a central place (hub) on-platform to provide end-user guidance on process, steps, associated time-frames.

It should be accountable, flexible and agile depending on the objective (purpose and scope) of reporting, for example aggregate reporting at set time periods, or a mix of both responding

to each and every case, and aggregate reporting, however as it concerns the latter criteria and parameters should be clearly set and outlined.

Full alignment where possible.

Additionally, the Irish Safer Internet Centre partner NPC conducted a survey aimed at both parents and children and young people for the purpose of this Call for input, and the findings revealed:

“Whilst 79% of parents said they were aware of being able to report content of concern to VSPS, only 48% had actually done so. 22% of parents were told of the outcome but only 9% were happy with the outcome. Many parents stated that they weren’t sure of the outcome as they had blocked the content or simply didn’t want to go back and check if it had been removed as it was just too distressing to view again.

“65% of young people were aware that they could report unsuitable content, and 43% had actually done so, but only 5% had been told of the outcome.

(NPC Survey 2023)

Question 10: What requirements should the Code include about age verification and age assurance? What sort of content should be shown by default to users who are logged out or in private browsing mode and whose age cannot be verified or assured? What evidence is there about the effectiveness of age estimation techniques? What current practices do you regard as best practice? Where accounts are not age verified should default privacy settings be used, should content default to universal content and should contact by others be more limited?

The practice of blanket age self-declaration to access products and services aimed at and/or used by minors must be ended. It is too easy for such a regime to be misused and by extension minors are not able to avail themselves of any protections a minor account would offer. Any method of age verification for the purposes of establishing whether a user is a child ought to be proportionate and risk-based. The Irish Safer Internet Centre endorses the CO:RE classification of risk as referenced earlier in this call (Q.2). Any method of age verification and age assurance must be privacy preserving and adhere to the principle of data minimisation.

Coimisiún na Meán in developing this code must adopt a similar approach as the Data Protection Commission demanding that such providers ‘...go the extra mile in proving that its measures around age verification...are effective’.¹⁹

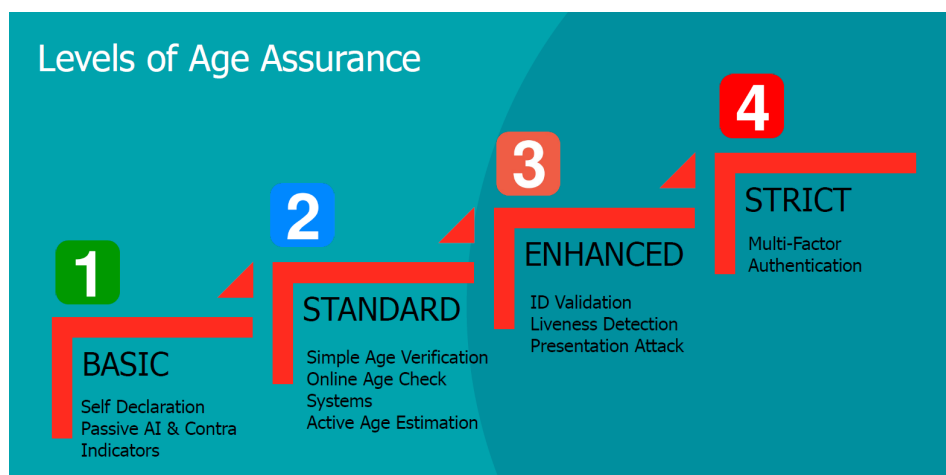
¹⁹ https://www.dataprotection.ie/sites/default/files/uploads/2021-12/Fundamentals%20for%20a%20Child-Oriented%20Approach%20to%20Data%20Processing_FINAL_EN.pdf

Attention must also be given to the work of the European Commissions' special group on the code of conduct on age-appropriate design.²⁰ This is a key action under the BIK+ Strategy. The code aims to reinforce the involvement of industry in protecting children when using digital products, with the ultimate goal of ensuring their privacy, safety and security online. It will be important to consider the direction this work is taking.

5 Rights Foundation in the UK list 11 common standards:for age assurance in its 'How do they know it is a Child' paper²¹:

- Age assurance must be privacy preserving
- Age assurance should be proportionate to risk and purpose
- Age assurance should be easy for children to use
- Age assurance must enhance children's experiences, not merely restrict them
- Age assurance providers must offer a high level of security
- Age assurance providers must offer routes to challenge and redress
- Age assurance must be accessible and inclusive
- Age assurance must be transparent and accountable
- Age assurance should anticipate that children don't always tell the truth
- Age assurance must adhere to agreed standards
- Age assurance must be rights-respecting

The Age Verification Providers Association on its website states that 'The Age Verification sector currently works to BSI PAS 1296:2018, and two more international standards are under development with the IEEE and ISO.²² It also speaks to levels of age assurance as the graphic below outlines. Levels of age assurance ought to be an important consideration when considering how to balance a child's rights to participate, their right to freedom of expression and their right to access information with their right to be protected.



²⁰ [The special group on the code of conduct on age-appropriate design](#)

²¹ https://5rightsfoundation.com/uploads/But_How_Do_They_Know_It_is_a_Child.pdf Page 5

²² <https://avpassociation.com/standards-for-age-verification/>

A submission on behalf of the Irish Safer Internet Centre to inform a future consultation by Coimisiún na Meán on a draft Online Safety Code

In the [euCONSENT project](#)²³, it is argued that a child-rights approach to age assurance must protect not only a child's right to be protected from digital content and services that could harm them, but also their right to privacy and freedom of expression (including to explore their identity or seek confidential help without parental consent), their right to a prompt and effective child-friendly remedy, and their right to non-discrimination. This means they must be able to access digital services along with everyone else even if they lack government ID or live in alternative care or have a disability, and whatever the colour of their face. At present, many systems of age assurance do not respect the full range of children's rights.

We recognise that the proposed European Commission funded euConsent project will not be live for another 12-18 months. This will be an EU-wide computer network for completing online age verification and securing parental consent when younger children wish to share personal data. The aim of this ground-breaking network is to protect children from harm on the web, particularly age-restricted goods, content and services while promoting their rights to the opportunities the internet offers.

The core principles are:

- Children have the right to participate in a digital world to the fullest extent possible.
- Providers of digital services and content directed at children should have a robust, trusted framework to deliver high quality age appropriate materials.
- People with parental responsibility or guardianship of children should have confidence in the standards and framework to enable permissive content for their children.
- Adult services and content should not be available to children to access (intentionally or by accident), and illegal content should not be tolerated.
- The regulatory eco-system should encourage market solutions through a robust framework of accreditation, certification and interoperability across the European Union.

The code ought to develop synergies with what is proposed in the euConsent project as any age verification scheme will also have to recognise that children aged 13-16 years depending on Member States need parental consent to share personal data on the basis of consent with VSPS, as per the General Data Protection Regulation (GDPR).

Evidence of effectiveness of age estimation techniques

Yoti has written a white paper on the effectiveness of its age estimation techniques.²⁴ It is reporting high true positive rates, including for minors. YOTI is a recognised model used by child protection agencies such as NSPCC and IWF²⁵. Reference to industry products is not an endorsement on behalf of the Irish Safer Internet Centre.

The ICO and Ofcom in the UK have also published a technical study on age assurance

²³ <https://www.lse.ac.uk/media-and-communications/research/research-projects/eu-consent>

²⁴ <https://www.yoti.com/blog/yoti-age-estimation-white-paper/>

²⁵ <https://www.yoti.com/blog/age-verification-childline-iwf-report-remove-nsppc/>

technologies.²⁶

Webwise Youth Panel Insights

“Younger children can easily bypass the 13+ or 16+ age requirement and access content that was not intended for them to access. Loot Crates and Loot boxes should be banned entirely as they do not have any guidance to what can be inside them and can cause a gambling problem from a young age.”

“I think they should make each app and video for people 13 and up and have to show some id to prove they are that age and are not lying about it.”

“Having a reliable way of showing that a person has parental consent, whether that is from a source of ID and a parent holding it or some other form of that would probably be helpful. Also having more ways of filtering through content, whether that be comments and reviews or videos and shows for parents that are simple and quick (seeing as parents cannot watch or view everything that their child does before they allow them to)..”

“Enforcing age requirements or some sort of Identification system in order to create responsibility for your actions and ensure you're mature enough to be in the online space”

“Having things that are not-age appropriate available for use by anyone. This could be in the form of (for example) websites or games being available with the click of a "I am above the age x" button, comments on a post that a person has placed online that could hurt, offend or upset them, or young people watching or reading etc. Things that they should be going through with their parents, but their parents are not available/able to monitor and discuss what they are doing online.”

On the topic of ‘How old do you think a child should be before they should be allowed to watch or share videos on websites or in apps? Should there be different rules for children who are different ages? Here’s what the Youth Panel had to say:

“I feel as though it is up to the individual family on how much they want their child to be exposed to content online. I feel as though starting secondary school was a good starting mark but it was been lowered drastically by society so now 7 or 8 year olds have access to inappropriate content.”

“Tricky. Kids like to watch cartoon videos etc... Having complete control over what they watch then maybe 13?”

“Different rules for children of different age ranges. I think the rules could be set by parents.”

²⁶ <https://www.drcf.org.uk/publications/papers/measurement-of-age-assurance-technologies>

“13 yes there should be different rules.”

“Different rules for different ages from ages 3 up.”

“I think it can depend a lot on how mature the kid is, and how educated the parents/child are when it comes to the internet. In general I think 4-9 yrs old should be monitored when watching videos, to make sure it's appropriate, as well as have a safe place to talk to parents about things, if you're a parent and you blame the child and yell at them for accidentally seeing a video they weren't meant to, that just promotes for them to never speak to you again on that matter. From age 10 to early teens, the person should be properly educated about the internet, like what to do if they find an inappropriate video (report/don't recommend button), and from middle teens (15-16) and late teens (16-19), they should hopefully be mature enough themselves to know what to do in such situations.”

“I think they should be a minimum age of 13 but young children (13-17) should have their videos more strictly monitored, both ones seen and posted.”

“13 to be allowed to share videos.”

“A child should be at least 12 or 13. No, the Internet should be for everyone. If they are allowed to watch it they shouldn't be sheltered.”

NPC's survey asked parents what types of age ratings (if any) should be applied for different video content, and

“the majority believed that there should be an age rating applied to most video content. Parents stated that adult, controversial and opinionated content should have an Age Assurance method to ascertain the age of the viewer, and a third of parents believed that fashion, beauty, personal development and lifestyle should have an Age Estimation method. Over a third of parents said that educational content such as DIY, cooking, fitness, sport, pets, and technology should only require Age Gating requirement.

33% of the young people who responded to the NPC survey felt it should be an official document, but interestingly, 24% said it should be an Age Gating method and another 24% said they should not be required to give their age.²⁷

(NPC Survey 2023)

²⁷ See appendices 2,3 & 4

Question 11: What requirements should the Code have in relation to content rating? What do you consider to be current best practice? What experiences have you had using content rating systems on platforms and do you think they have been effective? What steps could we ask VSPS to take to ensure content is rated accurately by users?

What do you consider to be current best practice?

PEGI

For nearly 20 years now, [PEGI](#) (Pan European Game Information) has been a successful example of self-regulation by providing advice to parents through age classifications. The PEGI age categories and content descriptions are designed to be simple and universally understandable. They are specifically designed for non-linear media and have been updated following technological, academic and societal developments. PEGI has a legal status across much of Europe with the notable exception of Germany where there is a separate system (USK.de).

The PEGI system covers console games, VR games, mobile and tablet games, and PC and cloud gaming. The notable exceptions are Apple and Steam which do not apply the PEGI system to their platforms and products. Most platforms now have good parental control tools. PEGI employs a code of conduct which is a set of rules to which every publisher using the PEGI system is contractually committed. The code deals with age labelling, promotion, and marketing, and reflects the video games industry's commitment to provide information to the public in a responsible manner.

What experiences have you had using content rating systems on platforms and do you think they have been effective?

In relation to content ratings, data from the National Parents Council 2023 survey revealed:

“

- 55% of parents said they were somewhat familiar with the content rating of video content and 54% favoured a system of age rating similar to that used for cinema content as a way of ascertaining whether content was suitable for their child or not.
- Many parents commented that they used social media sites for parents to verify whether content was suitable for their child.
- 48% of parents were not aware of any content rating information for selecting content on video sharing platforms, and 30% said they had only used them occasionally.
- 67% of parents felt that video sharing platforms did not provide enough information about their content to allow users to make informed decisions before watching them.
- 40% of the young people said they found descriptions of the content the most useful when deciding whether to view it or not, and 39% said the age ratings were more

A submission on behalf of the Irish Safer Internet Centre to inform a future consultation by Coimisiún na Meán on a draft Online Safety Code

useful, however, a majority of them (69%) said they were unaware or unsure if they had seen any of the platforms with these descriptions on them. 57% if they had seen the descriptions they may have changed their mind about viewing, and 47% said there was not enough information provided by the platforms before they viewed the content.”

(NPC survey 2023)

Webwise Youth Panel Insights:

“I would just like to stress the age verification when it comes to videos online. I feel as though young people can too easily come across inappropriate content on certain apps and it shouldn’t be allowed.”

“I’m concerned that some videos aren’t age checked before appearing on the feed. For example, I feel that oftentimes young teenagers come across content on the likes of tiktok that isn’t appropriate for their age.”

Question 12: What requirements should the Code have in relation to parental control features? How can we ensure that VSPS providers introduce the mechanism in a user-friendly and transparent way? Can you point to any existing example of best practice in this area? Should parental controls be ‘turned-on’ by default for accounts of minors or where age is not verified?

The code should set out a minimum standard in relation to parental controls that ensures ease of accessibility and takes a safety by design approach; i.e ‘turned on’ by default.

Taking into consideration the views of parents; research conducted by Vodafone with 750 parents of children aged 4 – 14, 88% of Irish parents worry about the content their children could see online.²⁸ Parents and carers need VSPS providers and similar services to play their part in protecting children online.

Offering parental controls is generally favoured but according to research such controls can give ‘a false sense of security’ and ‘not necessarily limit the online risk of harm’.²⁹

However, they can still have some role in child safety online. The code ought to require that VSPS providers offer a suite of parental controls to parents and carers with the recommendation to involve their child and young person in any conversations on the use of

²⁸ <https://www.ispcc.ie/88-of-irish-parents-worry-about-the-content-their-children-could-see-online/>

²⁹ <https://euconsent.eu/download/understanding-of-user-needs-and-problems-a-rapid-evidence-review-of-age-assurance-and-parental-controls/>

parental controls. Industry should also be asked to consult with children and young people on what parental control features they feel work well. Ultimately, the safety of a product or service is down to the provider and not parents/carers.

Any controls must respect, protect and fulfil children's rights; be accessible; turned on by default; easily navigable; recognising and accommodating to children's age and stage of development (i.e. child's evolving capacity).

The NPC 2023 survey found that

“the vast majority of parents were aware or at least somewhat aware of parental controls that are available on digital devices and online platforms; (95%), with 49% of them using them regularly, 33% using them occasionally and 17% not using them at all.

Only 13% of parents were confident in their ability to use parental control features to manage the content their children could access and 10% of parents did not feel confident at all. 94% of parents thought that parental controls should be turned on by default.

Only 35% of young people were aware of parental controls, but 51% felt they should be turned on by default.”

How can we ensure that VSPS providers introduce the mechanism in a user-friendly and transparent way?

By introducing minimum standards that are accessible and developed in consultation with parents/parental organisations and children and young people.

Question 13: What requirements should the Code contain to ensure that VSPS provide for effective media literacy measures and tools?

The Irish Safer Internet Centre believes media literacy is a crucial skill for all ages and given online sources and social media are being used more frequently as the main source of news in Ireland, particularly among younger people; media literacy tools and education is more important than ever. Clear requirements should be provided for in the code for effective media literacy measures that also raise users' awareness of those measures/tools and fully align to the DSA.

According to a [report](#) analysing the current state of digital news in Ireland, it found for people aged between 18 to 24 years, nearly 40% of people chose social media as their main news source. The report also highlights worries about misinformation and disinformation are

A submission on behalf of the Irish Safer Internet Centre to inform a future consultation by Coimisiún na Meán on a draft Online Safety Code

growing. Concern about what is real and what is fake on the internet is comparatively high in Ireland (64%), up 6pp since last year. Furthermore, concern over fake news has increased over the past year in all territories surveyed: up 6pp in Ireland, 8pp in the UK, 4pp in the US, and 2pp in Europe. UK news consumers are the most concerned about fake news and misinformation online, with 69% saying the issue is concerning. This across-the-board increase over the past year is perhaps connected to fears that news content - especially so-called 'deep fake' photos and videos - is being produced by AI technology.³⁰

An [Ipsos Mori survey from March 2021](#) found that just 9% of Europeans (from 11 countries) have participated in training about how to use online tools to distinguish between true and false information, but 58% are interested in doing so. Two-thirds of those surveyed believed it would be appropriate for a tech company to provide training to users to improve their ability to critically understand online information.³¹

Digital literacy is essential for supporting children's growth online, the 2023 Ofcom Children and Parents: Media Use and Attitudes report highlights the need for a continuous focus on digital literacy to combat negative feelings and misinformation on social media, as a third of children 'believed all or most of what they saw on social media to be accurate and true.'³² Furthermore a recent paper; Digital literacy and online resilience as facilitators of young people's wellbeing? A systematic review notes; Digital literacy functions as a promotive factor of wellbeing, providing beneficial outcomes in different areas of life but also shielding young people from harm as a result of online risk experiences.³³

It is relatively easy to create and disseminate dis/misinformation online which can reach wide audiences through the amplification of algorithms that have little transparency or information on where content has come from or who it has been created by. To address the issue, it's crucial to look at the business models tied to the ad tech industry, enhance the capability to identify disinformation and fraudulent online behaviour, and help users in critically assessing content. Social media platforms have a major role to play in assisting with this learning process, monitoring behaviour is not enough. Measures and tools need to be accessible and users made aware of the tools for example through the use of prompts/nudges.

Webwise Youth Panel Insights - during consultations and discussions with the Webwise Youth Advisory Panel in 2020, one teen noted:

³⁰ https://www.cnam.ie/wp-content/uploads/2023/06/20230609_DNR-Final-Report_STRICT-EMBARGO-00.01-14-June-23_FINAL.pdf

³¹ <https://www.ipsos.com/en-uk/online-media-literacy-across-world-demand-training-going-unmet>

³² <https://www.ofcom.org.uk/research-and-data/media-literacy-research/childrens>

³³ Vissenberg, J., d'Haenens, L., & Livingstone, S. (2022). Digital literacy and online resilience as facilitators of young people's wellbeing? A systematic review. *European Psychologist*, 27(2), 76-85. doi:10.1027/1016-9040/a000478

“I think that the [OSMR] bill should hold companies accountable for the spreading of fake news because I genuinely believe that it is one of the most predominant issues currently and the possibility of a post being fake news needs to be outlined to the user.”

Webwise Youth Panel Member, Aged 16

“In the social media platform ‘Twitter’, fake news is highlighted with an icon which appears above the post notifying the user ‘this information could be false’ or similar. I believe that all platforms should have an icon letting the user know that the information could be false. It would help stop the spread of fake news. Recognising faux news should also be included in school curriculum.”

Webwise Youth Panel Member, Aged 16

Question 14: How should we ask VSPS providers to address online harms in their terms and conditions in the Code, including the harms addressed under Article 28b? How should key aspects of terms and conditions be brought to users’ attention? What examples are there of best practice in relation to terms and conditions including content moderation policies and guidelines?

Clear requirements should be provided for in the code for VSPS ensuring a minimum standard of clear terms and conditions for their content and content moderation practices.

Webwise Youth Panel Insights - when asked about terms and conditions, members of the Webwise Youth Panel have been vocal about the need for clear, accessible information. Here is what they had to say on the matter:

“The Terms and Conditions in apps should be simpler and more accessible to read in a way that is visually pleasing and gathers the attention of the reader. This design also should include simple terms for younger users of social media (13 and over) with shorter main descriptions, focusing on how their data is used...”

Webwise Youth Panel Member, Aged 16

“Bringing in better and more clear guidelines for behaviour on social media and implement more accurate fact checking methods”

“One issue I find is that social media companies don’t explain to their users in detail the facilities that they have as so many young people aren’t aware of the support systems are in place, this could be changed by having informational videos or posts to outline their supports and how to use them effectively. Reporting needs to be better regulated and more efficient as I have noticed many times that reporting goes unnoticed and nothing gets done, which is a big problem.”

A submission on behalf of the Irish Safer Internet Centre to inform a future consultation by Coimisiún na Meán on a draft Online Safety Code

Webwise Youth Panel Member, Aged 16

“...explainer videos for different functions on apps around safety and privacy settings and reporting, I think that’s a great idea. Also if social media apps were to have their settings be laid out in a specific way, or if there was a clearer direction to your privacy and safety settings, as it is often extremely hard to change certain things or even find settings in some apps.”

Webwise Youth Panel Member, Aged 18

“While super-reporting is a great measure to ensure that the reporting process is not always extremely lengthy and exhaustive, it may detract from the voices and opinions of young people who may not be able to get their view to an organisation that is a part of the process, if it was to split the reporting process so that there is a normal process and a fast-track process, while the fast-track process is useful, it is neither helpful nor favoured by people who are participating in the normal process, and could slow their experience even more. With the implementation of this, it would be good to focus on making sure youth voices and the voices of ordinary people in general don’t get drowned out.”

Webwise Youth Panel Member (Aged 18)

“Perhaps an easier way to report false information on apps and social media and a faster action by social media companies to remove this false information/false accounts/misleading posts.”

“I believe it is extremely important as people don’t understand what they could be agreeing too. These terms & conditions are too wordy and may be difficult especially for visual learners and people with reading difficulties. This problem doesn’t just affect people’s ability to understand it affects their personal data.”

“...I think easier and clearer terms and conditions are a must. Younger audiences won’t want to scroll through endless amounts of small print that may be challenging to understand. I think bright colours, imagery, audio/video and simpler wording is vital to this. I also think there should be questions to be answered at the end in order to prove that the terms and conditions have been acknowledged, not just clicked through. I also think there should be a verification of age to prevent those underage giving a fake date and being allowed access to online services and platforms. This is done by the likes of unidays and spotify students when verifying they are a student by providing an image of their student ID which is reviewed before being verified to have access to the account.”

Webwise Youth Panel Member, Aged 20

“The biggest problem with being online is that youths don’t understand certain aspects of being online and what it entails such as: terms and conditions being difficult to understand, what does it mean for a website to take cookies or what am I agreeing to gain access to this

website. Young people don't understand the sensitivity of some of the actions and what it does to their data. So the biggest problem we are facing is children and adults alike not understanding what they are agreeing to."

"I believe it is extremely important as people don't understand what they could be agreeing to. These terms & conditions are too wordy and may be difficult especially for visual learners and people with reading difficulties. This problem doesn't just affect people's ability to understand, it affects their personal data."

"They must make it in law that websites and social media alike must have a visual option for terms and conditions that is easy to comprehend."

For members of the Webwise Youth Advisory Panel this is an important issue and one that would benefit from proper consultation with young people and other vulnerable groups.

Best practice in relation to terms and conditions including content moderation policies and guidelines

The Fundamentals for a Child-Oriented Approach to Data Processing³⁴

Question 15: How should we ask VSPS providers to address content moderation in the Code? Are there any current practices which you consider to be best practice? How should we address automated content detection and moderation in the Code?

Content moderation is multipronged, and the Code should address and include minimum standard requirements for VSPS.

As a first port of call, risk and impact assessment informed processes tailored to the nature of harm and content type being moderated as one solution does not fit all, however there would be baseline common denominators. As such the fundamental risk and impact assessment criteria could be prescribed, however allowing flexibility to VSPS on conducting the risk and impact assessment within the full breadth of technical specificities of their service and potential emerging trends associated with continuous development of the service and the users' use of the same. The outcome of the risk and impact assessment would become the blueprint informing the development of bespoke processes, subsequently identifying the most adequate vehicle for content moderation (automated, tech-enabled and human moderation or oversight).

Additionally, there is an overdue need to establish content moderation industry standards e.g.

³⁴ <https://www.dataprotection.ie/en/dpc-guidance/fundamentals-child-oriented-approach-data-processing>

specialised training, must-have skill-sets and expertise, staff-welfare and support, quality assurance, to name a few.

Question 16: What requirements should the Code include about procedures for complaint-handling and resolution, including out-of-court redress or alternative-dispute resolution processes? To what extent should these requirements align with similar requirements in the DSA? What current practices could be regarded as best practice? How frequently should VSPS providers be obliged to report to the Commission on their complaint handling systems and what should those reports contain? Should there be a maximum time-period for VSPS providers to handle user complaints and if so, what should that period be?

Clear requirements about complaint handling and resolution should be provided for in the code.

For effective reporting mechanisms the following factors should be recommended for platforms:

- Ease of reporting: Reporting mechanisms should be easy to access and use and designed with all internet users in mind, particularly vulnerable users. A consistent approach to reporting should be advised across platforms.
- Ensuring reporting for non-account holders
- Clear routes for redress

Education

Education about what to report: There is the need for education for users about what is not OK online, and what people should report, as well as that it is worth making a report about harmful content.

Education about how reporting works: This would include being clear about the anonymous aspects of reporting for example, education on how best to report, and identifying the barriers young people face, and taking steps to overcome these. Education will have an online and an offline component.

Support reporting offline: it is important to think about the offline support that can be offered, and this includes in the instances where the content is not taken down. Signposting to other related organisations that can help, for example, Childline by ISPCC.

Listen to users on reporting: Regular research with users, particularly examining awareness and confidence in reporting mechanisms, will help to inform both the Codes and the additional needs in this area, recognising that these needs may fall outside of industry reporting.

Training for professionals working with children: as a high degree of issues are reported via schools and established child protection and safeguarding processes, those working with them need to be able to access training to ensure competency in their ability to recognise, respond to, and resolve issues related to online harms with under 18s.

It is vital that users are made aware of what's available to them, what they can report and what their rights are. Information on these important changes and the rights of users should also be communicated in clear, accessible, multi-format manner by the Online Safety Commissioner as well as the VSPS.

Webwise Youth Panel Insights:

On whether you ever reported your concerns to your parent/s or guardian/s or to a company in charge of websites or apps about a video that you have seen? How did that go... 9 students indicated they have never taken this action. Here's what the Webwise Youth Panel had to say:

"I have not reported content to anyone in particular. I report it on the app and block the publisher."

"Personally I have a very open relationship with my parents about content I see online but I know many young people do not have such a relationship. As for the websites/apps, I feel I have often reported accounts or videos and felt that it just never gets sorted out and that the companies rarely do anything about."

When asked about the extra support people would need to step in and defend the targets of online bullying, respondents to the recent [Bystander Behaviour Online Among Young People in Ireland](#); Most participants suggested implementing some kind of technical improvement or a better management from the social media or digital service providers, with several participants calling for the facilitation of reporting and be provided a prompt response to the situation.

"A quicker response from social media platform when you report someone or something" (Boy, 5th Year)

"There should be a button to leave anonymous reviews about them to the online app and then they can handle it from there" (Girl, 3rd Year)

"Word blocker" (Boy, 4th Year)

The report highlights social media providers *can* contribute to reducing cyberbullying. The mechanics of some social networking sites could be facilitating online victimisation given the higher rates found in this study of witnessing cyberbullying among those registered in some social networks in particular. This requires further research before conclusive recommendations can be made, but several students themselves called for technical improvements on social media and engagement from the providers to facilitate other people

stepping in when encountering cyberbullying

What current practices could be regarded as best practice?

- National human rights institutions (NHRIs) Series: Tools to support child-friendly practices. Child-Friendly Complaint Mechanisms (P.62-63):
https://www.unicef.org/eca/sites/unicef.org.eca/files/2019-02/NHRI_ComplaintMechanisms.pdf
- Handbook for policy makers on the rights of the child in the digital environment:
<https://rm.coe.int/publication-it-handbook-for-policy-makers-final-eng/1680a069f8>
- Children’s Rights and Business Principles
<https://www.unicef.org/documents/childrens-rights-and-business-principles>
- Council of Europe Guidelines to respect, protect and fulfil the rights of the child in the digital environment
[Guidelines to respect, protect and fulfil the rights of the child in the digital environment](https://rm.coe.int/guidelines-to-respect-protect-and-fulfil-the-rights-of-the-child-in-the-digital-environment/1680a069f8)

Question 17: What approach do you think the Code should take to ensure that the safety measures we ask VSPS providers to take are accessible to people with disabilities?

The Irish Safer Internet Centre believes accessibility should be a ‘must-have’, on a par with privacy, security and safety by design.

There should be a clear requirement for accessibility to be built by design and co-created in consultation with expert bodies such as the National Disability Authority (NDA) whose work is guided by the United Nations Convention on the Rights of Persons with Disabilities. It also incorporates the Centre for Excellence in Universal Design (CEUD), which is the only statutory Centre of its kind in the world.

There remains a scarcity of information about the experiences of children with disabilities. To address this gap the Council of Europe commissioned a study to explore the children’s views on how their rights were realised in relation to: access to the digital environment; impact on education, health, play and recreation; safety and protection; opportunities for increasing involvement in decision-making. The research Two Clicks Forward and One Click Back, Report on children with disabilities in the digital environment³⁵ notes “the challenges and barriers faced by children with disabilities vary significantly according to the type and nature of the impairment. It does them a disservice to lump them together as an undifferentiated group”. [...] “It was apparent throughout the study that laws, policies and services on the digital environment, that conflate children of different ages, living in different contexts and

³⁵ <https://rm.coe.int/two-clicks-forward-and-one-click-back-report-on-children-with-disabili/168098bd0f>

A submission on behalf of the Irish Safer Internet Centre to inform a future consultation by Coimisiún na Meán on a draft Online Safety Code

with different disabilities under the single heading ‘children with disabilities’, have the potential to do them a disservice, underplaying the significant diversity in their lived realities of the digital world.”

It further reveals, “While some of the challenges faced do not have digital solutions, technological developments have enabled many children with disabilities to find information, communicate, socialise, learn and play in ways that were not previously possible or are still not possible to the same extent in their non-digital lives.”

Safety measures and complaints mechanisms and solutions should be appropriately tailored, clear and accessible to all users regardless of age, ability, or disability.

A Webwise Youth Panellist noted;

“I believe it is extremely important as people don't understand what they could be agreeing too. These terms & conditions are too wordy and may be difficult especially for visual learners and people with reading difficulties. This problem doesn't just affect people's ability to understand, it affects their personal data.”

Question 18: What approach do you think the Code should take to risk assessments and safety by design? Are there any examples you can point us towards which you consider to be best practice?

Child Rights by Design

In the Digital Futures Commission, with 5 Rights Foundation, have proposed a model of Child Rights by Design; a principled vision to inspire innovators to help realise children’s rights when designing digital products and services. It provides a toolkit for designers and developers of digital products and was co-developed with them – and with children. It draws on the UNCRC and General Comment 25³⁶. It centres on 11 principles of which age-appropriate service is one, privacy is another, also safety, of course. The other eight are equally important for a holistic approach – equity and diversity; best interests; consultation with children; business responsibility; child participation; wellbeing; fullest development; and agency in a commercial world³⁷.

The Australian eSafety Commissioner’s [Safety by Design](#) provides a model for industry of all sizes and stages of maturity, providing guidance as they incorporate, assess and enhance user safety. The safety principle approaches online risks and harms from the social dimension of technology use. The approach focuses on embedding safety into the culture and leadership of an organisation. It emphasises accountability and aims to foster more positive, civil and

³⁶ <https://www.ohchr.org/en/documents/general-comments-and-recommendations/general-comment-no-25-2021-childrens-rights-relation>

³⁷ <https://inform.org/2023/07/20/how-can-we-make-the-internet-safe-for-children-in-practice-sonia-livingstone/>

rewarding online experiences for everyone.

Safety by Design Principles

1. Service provider responsibility

The burden of safety should never fall solely upon the user.

2. User empowerment and autonomy

The dignity of users is of central importance.

3. Transparency and accountability

Transparency and accountability are hallmarks of a robust approach to safety.

Webwise Youth Panel Insights

On the subject of whether companies who run websites or apps that allow videos to be watched or shared should do anything to make things safer for you or your friends or family, here's what the Webwise Youth Panel had to say...

"heavier restrictions and keeping notice on ways people are getting around those restrictions such as censoring words with numbers."

"Don't allow harmful or offensive videos to be posted"

"Age restriction, parent passcode, a warning before videos starts"

"I think it's the companies responsibility to ensure their products are safe to use thus I agree that they should make things safer."

"I think it's there website and if I don't like it that's on me"

"Yes, especially for younger kids, I've heard some stories that some things get past to YouTube kids because the algorithm isn't the best, as well as videos on YouTube intended for a more mature audience (13+ or 16+) get flagged for kids if they don't have swearing or violence in it, which leads to some problems for the creators themselves."

I think that the videos that appear on feeds / algorithms should be better programmed to age.

"No they already have child safety features like parental control"

..."Loot Crates and Loot boxes should be banned entirely as they do not have any guidance to what can be inside them and can cause a gambling problem from a young age."

"Help tech companies put in better way to verify your age and ban loot crates"

A submission on behalf of the Irish Safer Internet Centre to inform a future consultation by Coimisiún na Meán on a draft Online Safety Code

“EU policy makers could help by forcing all platforms to have privacy settings. Things like limiting comments to certain people (e.g. only people you follow back on social media) and being able to have a private account can make a huge difference in protecting young people and children online.” Webwise Youth Panel Member.”

Question 19: How do you think that cooperation with other regulators and bodies can help us to implement the Code for VSPS?

Collaborative cross-border partnerships and peer advisory and support initiatives such as the Global Online Safety Regulators Network are invaluable and a forum that would hopefully enable the development of global gold standards of governance, regulation, policy, and practice for online safety. Harnessing the power of cross-nations and borders knowledge whilst in the unique position to have first-hand insights into both common denominators and differences would be necessary in tackling and reducing harms manifested on a global scale with the potential of impacting anyone’s life at any time and having long lasting consequences.

There will also be a need to cooperate, for example, with the Data Protection Commission for GDPR compliance and other matters such as age verification and age assurance mechanisms and approaches for the purpose of knowing the age of minors using such products and services.

Cooperation with organisations who will act as trusted flaggers and who will support the delivery of the super complaints scheme will also be important.

Question 20: What approach do you think we should take in the Code to address feeds which cause harm because of the aggregate impact of the content they provide access to? Are there current practices which you consider to be best practice in this regard?

There is growing evidence on the harmful consequences of algorithms used by large online platforms. According to the office of the [eSafety Commissioner Australia](#); the question of whether content served up by a recommender system is harmful can depend on the individual user, their personal circumstances and the context.

For example, content that promotes self-harm is likely to present a greater risk and have a deeper impact for someone already experiencing mental ill health. In addition, the risks can be greater for children and young people, especially if they are served:

- friend or follower suggestions that encourage them to interact with potentially dangerous adults
- content that encourages binge consumption without breaks
- content that promotes ‘ideals’ of body types and beauty stereotypes

A submission on behalf of the Irish Safer Internet Centre to inform a future consultation by Coimisiún na Meán on a draft Online Safety Code

- content that normalises the sexualisation of young people
- content that may be appropriate for adults but harmful to children who are not developmentally ready for it.

Recommender systems also have the potential to cause or worsen harms on a societal level. For example, content that promotes discrimination such as sexism, misogyny, homophobia or racism can normalise prejudice and hate. It can also be used to incite online pile-ons or physical violence that can cause damage to the people targeted and spill over to affect the broader community, both online and offline³⁸.

Pathways: How digital design puts children at risk

The [Pathways report](#) is the outcome of a research project undertaken by Revealing Reality on behalf of 5Rights Foundation. It examines how design choices embedded in digital products impact the lives of children. Through interviews with digital designers and children, and through innovative research using avatars, it lays bare how the commercial objectives of digital companies translate into design features that impact on children³⁹.

It should also be noted that recommender systems and algorithms have many benefits to users for example; more targeted search results, help users discover new information, and give a more personalised experience online. What is displayed on news feeds or search results is determined by the algorithm of the platform a person is using, and while it is based on a number of things such as personal interests and how engaging the content is, the exact details of why and how they work are largely unknown. See also: 5Rights Foundation Disrupted Childhood 2023⁴⁰.

According to Esme Fowler-Mason: “Algorithms also amplify extreme content because this is what keeps us engaged. Whilst this can be positive, it also fosters the growth of [paedophile rings on YouTube](#), [extreme right-wing groups on Facebook](#), and [pro-eating disorder communities on TikTok](#).”⁴¹

Algorithmic Accountability and Transparency

The Digital Services Act contains several obligations, with the goal of increasing algorithmic transparency and accountability. Any such measures included in the Online Safety Codes need to align to the requirements of the DSA.

³⁸ <https://www.esafety.gov.au/industry/tech-trends-and-challenges/recommender-systems-and-algorithms/full-position-statement>

³⁹ <https://5rightsfoundation.com/uploads/Pathways-how-digital-design-puts-children-at-risk.pdf>

⁴⁰ <https://5rightsfoundation.com/uploads/Disrupted-Childhood-2023-v2.pdf>

⁴¹ <https://blogs.lse.ac.uk/medialse/2023/02/08/the-online-safety-bill-needs-more-algorithmic-accountability-to-make-social-media-safe/#:~:text=The%20term%20'harm'%20covers%20a,keeping%20us%20online%20for%20longer> .

Are there current practices which you consider to be best practice in this regard?

- European Centre for Algorithmic Transparency
https://algorithmic-transparency.ec.europa.eu/index_en
- UNICEF's Policy Guidance on AI for Children:
<https://www.unicef.org/globalinsight/reports/policy-guidance-ai-children>
- 5Rights: How digital design puts children at risk
<https://5rightsfoundation.com/uploads/Pathways-how-digital-design-puts-children-at-risk.pdf>

Webwise Youth Panel Insights - there were mixed views from the Webwise Youth Advisory Panel on whether they have enough control over the type of videos that you see on websites or apps:

“When you report a post on most websites it restricts the content on your feed, meaning the more you ignore or do not interact with a post, the less that topic will come up. I report videos if they do not seem appropriate or seem to be harmful.”

“To a certain extent, I don't have to watch the videos I can leave them if I choose”

“No, I don't think there is enough control.”

“Most of the time, however, if I see anything I dislike or I know is wrong I report it and block whatever it is.”

“Yup, since I can block content creators whose content I don't like/is inappropriate as well as click 'don't recommend' button when watching shorts so they don't pop up again.”

“On some apps yes but on tiktok of Instagram reels I feel that all control is lost to the algorithm.”

“No you could be recommended anything”

In addition, the 2023 NPC survey found that whilst 72% of parents surveyed were well aware of content feed and how it works, a third of young people surveyed had no knowledge of content feed.

Question 22: What compliance monitoring and reporting arrangements should we include in the Code?

The Irish Safer Internet Centre supports the proposal in this call document to require VSPS providers to provide an annual compliance statement, approved by the Board of Directors of a VSPS provider as a form of compliance monitoring. It would be useful that such documents, and/or an appropriate version of the same (any commercial sensitivities removed) are made public to support confidence in products and services.

A submission on behalf of the Irish Safer Internet Centre to inform a future consultation by Coimisiún na Meán on a draft Online Safety Code

It is also recommended to defer to The Fundamentals⁴² by the Data Protection Commission for direction on compliance monitoring and reporting where children are concerned.

In respect of reporting arrangements specifically, in defining the requirements it would be necessary to clearly identify the scope of the reporting, the audience and the level of dissemination (e.g. general public release, restricted e.g. for the purpose of informing regulatory compliance and insights).

To that end, as deemed appropriate within the purpose, scope, and audience, the Irish Safer Internet Centre recommends the following information, non-exhaustive and in no particular order, might provide practical insights:

- (i) the nature, context and content of the relevant material and the severity of its impact and harm;
- (ii) the extent of avenues available or suitable to address the type of harm and whether such approaches have been successful or not;
- (iii) harm reduction indicators and measurement;
- (iv) preventive and deterrent measures deployed and the effectiveness in harm reduction;
- (v) whether the intended subject of the regulatory action has been the subject of prior compliance or enforcement action, and the outcome of that action;
- (vi) the extent to which any conduct represents a broader systemic issue;
- (vii) the circumstances of the end-user any indicators of vulnerability and level of support required to respond to compliance or enforcement action;
- (viii) emerging trends and issues identified during the reporting period;
- (ix) the action times on complaints handling broken down per type of resolution.

⁴² https://www.dataprotection.ie/sites/default/files/uploads/2021-12/Fundamentals%20for%20a%20Child-Oriented%20Approach%20to%20Data%20Processing_FINAL_EN.pdf

Timestamp	How old are you?	Gender
-----------	------------------	--------

7/20/2023 17:09:27		17 female
--------------------	--	-----------

7/20/2023 17:09:48		15 Female
--------------------	--	-----------

7/20/2023 21:59:49		17 Female
--------------------	--	-----------

7/20/2023 23:03:39		17 Female
--------------------	--	-----------

7/22/2023 9:42:15		16 Male
-------------------	--	---------

7/24/2023 15:59:52		16 Male
--------------------	--	---------

7/24/2023 17:04:27

17 M

7/24/2023 20:54:32

16 Female

7/27/2023 8:36:09

19 Female

8/2/2023 22:30:12

17 Male

8/6/2023 10:22:12

16 Male

What do you like about being able to watch or share videos on websites or apps?

is a fun way to be educated on topics instead of reading books and research papers.

They are entertaining and I can share with family and friends

That's it's easy and handy. Only takes a few seconds to share.

I like being able to watch various topics that have an interest to me and learn new things about people, countries, cultures, etc. I like to be able to share videos to show friends and family whatever content it is. Being able to share videos also allows us to promote or spread awareness if it is a certain type of content that should be made known.

You can learn new things on these websites

It's a great way of learning new things

That it's easy to share information quickly

Being able to show things to my friends and find things that inspire me, as well as be able to learn about different cultures on YouTube. I also enjoy watching videos to find new music, art, and creators that I enjoy.

Feeling in touch with other people and the social aspect of it.

Entertainment

Information spreadsheet p
Fast

How safe do you feel when you are watching or sharing videos on websites or apps?

Are you concerned about any videos that you see on websites or on apps? If you are, what types of videos concern you the most?

I feel safe using apps such as YouTube, tiktok and Instagram as they are heavily monitored and have strict restrictions. Apps such as Twitter, tumblr and third party websites are less monitored and I find myself seeing offensive or inappropriate content more often.

pornography and hate speech is what I have been most exposed too.

Not 100% safe as we are not sure of who is watching our moves or tracking us

Yes, homophobic, sexist and racist videos.

Not too safe with websites and links. Especially getting links for videos or websites from people I don't know

Some videos on tiktok are not suitable for younger kids to watch. It worries me as I see how young some of the people on tiktok are.

Personally I feel safe as I'm aware of the dangers online and never venture into the unsafe aspects of being online.

I don't find myself concerned

I feel safe enough

No not really

Mostly safe

Yes. Videos of fights

Very safe

No

I feel rather safe on YouTube when it comes to sharing videos or watching videos, as it tends to be my main platform for watching videos.

Not particularly, I myself don't get shown videos that are inappropriate, but then again I am at an age where not too many things are very inappropriate, but I don't see content with sex or extreme gore pop up.

It depends on the website/app and how well educated user are about their safety features

I'm concerned that some videos aren't age checked before appearing on feed. For example, I feel that often times young teenagers come across content on the likes of tiktok that isn't appropriate for their age.

Quite safe

No

Quite safe and entertained

Not really

Do you feel that you have enough control over the type of videos that you see on websites or apps?

Do you think that companies who run websites or apps that allow videos to be watched or shared should do anything to make things safer for you or your friends or family?

when you report a post on most websites it restricts the content on your feed, meaning the more you ignore or do not interact with a post, the less that topic will come up. I report videos if they do not seem appropriate or seem to be harmful.

heavier restrictions and keeping notice on ways people are getting around those restrictions such as sensoring words with numbers.

To a certain extent, I don't have to watch the videos I can leave them if I choose

Don't allow harmful or offensive videos to be posted

No I don't think there is enough control.

Age restriction, parent passcode, a warning before videos starts

Most of the time, however if I see anything I dislike or I know is wrong I report it and block whatever it is.

I think it's the companies responsibility to ensure their products are safe to use thus I agree that they should make things safer.

I guess so

I think it's there website and if I don't like it that's on me

No

Yes

Yes

Maybe add better age verification

Yes, especially for younger kids, I've heard some stories that some things get past to YouTube kids because the algorithm isn't the best, as well as videos on YouTube intended for a more mature audience (13+ or 16+) get flagged for kids if they don't have swearing or violence in it, which leads to some problems for the creators themselves.

Yup, since I can block content creators who's content I don't like/is inappropriate as well as click 'dont recommend' button when watching shorts so they don't pop up again.

On some apps yes but on tiktok of Instagram reels I feel that all control is lost to the algorithm.

I think that the videos that appear on feeds / algorithms should be better programmed to age.

No

Yes

No you could be recommended anything

No they already have child safety features like parental control

How old do you think a child should be before they should be allowed to watch or share videos on websites or in apps? Should there be different rules for children who are different ages?	Have you ever reported your concerns to your parent/s or guardian/s or to a company in charge of websites or apps about a video that you have seen? How did that go?	Is there anything else you would like to comment on?
I feel as though it is up to the individual family on how much they want their child to be exposed to content online. I feel as though starting secondary school was a good starting mark but it was been lowered drastically by society so now 7 or 8 year olds have access to inappropriate content.	I have not reported content to anyone in particular. I report it on the app and block the publisher.	I think it is an excellent idea that young people are getting our chance to give our views and experiences online
Tricky. Kids like to watch cartoon videos etc... Having complete control over what they watch then maybe 13?	No	No
Different rules for children of different age ranges. I think the rules could be set by parents	N/a	No
I think depending on the app it should depend on the age restrictions.	I've never	No
13	No	No
yes there should be different rules	No	No

Different rules for different ages from ages 3 up

No

No

I think it can depend a lot on how mature the kid is, and how educated the parents/child are when it comes to the internet. In general I think 4-9 yrs old should be monitored when watching videos, to make sure it's appropriate, as well as have a safe place to talk to parents about things, if you're a parent and you blame the child and yell at them for accidentally seeing a video they weren't meant to, that just promotes for them to never speak to you again on that matter. From age 10 to early teens, the person should be properly educated about the internet, like what to do if they find an inappropriate video (report/don't recommend button), and from middle teens (15-16) and late teens (16-19), they should hopefully be mature enough themselves to know what to do in such situations.

No

Nope

every open relationship with my parents about content I see online but I know many young people do not have such a relationship. As for the websites/apps, I feel I have often reported accounts or videos and felt that it just never gets sorted out and that the

I would just like stress the age verification when it comes to videos online. I feel as though young people can too easily come across inappropriate content on certain apps and it shouldn't be allowed.

I think they should be a minimum age of 13 but young children (13-17) should be have their videos more strictly monitored both ones seen and posted.

13 to be allowed to share videos

No

No

A cgild should be at least 12 or 13. No the Internet should be for everyone if they are allowed watch ut they shouldn't be sheltered

No

No

Parent's Comments responses to question 21

1. Platforms should be legally responsible for ensuring that inappropriate content cannot be accessed by users based on their age. It is nonsense to think that parents have the technological mastery to be able to protect their children from harmful content. All adult material should require definitive age verification.
2. Why weren't the dangers of gender ideology and the promotion by radical trans activists of the castration of children listed as one of the dangers lurking online for children? This is more dangerous than the promotion of eating disorders online....
3. You tube has a kids version but it's too childish for my 10 year old. I have blocked explicit content on the adult version but I do worry about something unsuitable coming up. I would like to see a version more suitable for tweens/teens.
4. More awareness and advertising of the full consequences of what children can and are exposed to on social media needs to happen. The true extent of the harm it causes to the vulnerable or immature mind requires much more airtime. I don't believe the lay person fully understands the problems with social media themselves and the above platforms so therefore can't possibly look out for their child using any of the above platforms. France and even the UK seem to be ahead of Ireland in terms of the protection of children on these platforms. Also I believe the platforms themselves and content creators need to have a larger level of responsibility for the content and who can access it. At the end of the day these companies are making money from views as are the creators of the content so they need to be held accountable like any other business. The content creators solely care about views and clicks for advertising and income so who views it or makes the click is of no concern to them as the money is already been made. More awareness needs to be made around how much revenue the above companies make and the content creators and then people would know how valuable their child's time on these platforms is actually worth.
5. Questions that pop up like " what is 9 * 7 " stops a young and maybe naive person going further with posting
6. Stricter control on what advertisements are played before and during Youtube videos. Several times, when our young daughter was watching cartoons on Youtube, inappropriate ads were played. One was for a horror film.
7. Ban unsuitable content full stop. Why should it be allowed? It's immoral anyway. It's far too much pressure on parents to monitor unsuitable content. The platforms should not allow bad language sexual content or violence.
8. Try not to let underage children on .
9. I agree with banning the use of mobile phone devices for primary school children
10. Make it mandatory that ALL children with & without additional needs do not have access to inappropriate, harmful content
11. I wish there were more towns to follow the example of Greystones, where parents collectively decided they wouldn't let their children have a smartphone before they go to secondary school - and I wish I could live in one of them. Maybe prompting and encouraging such initiatives would be beneficial.
12. Make it safer
13. My children will not use social media or sharing platforms until they are 16. None of their friends do either (aged 10 -13) as their parents are on the same page.
14. Disable online comments preventing online bullying

15. Take off the pop ups and put more safety measures in place for them. Also have safety measures automatically on social media platforms (let adults turn them off) or make it easier for adults to put them on!
16. Children will always find ways around parental controls. The platforms need to be more responsible and face penalties otherwise. Porn can be viewed in harmless seeming apps like pinterest. Tiktok very quickly brings young people through to videos on disordered eating and upsetting emotional content. Youtube kids is a good more though not perfect, but doesn't have enough content for a tween, same as Spotify. They should all restrict content based on age. However, I don't want to submit my child's passport or have their face scanned! I would be happy to set up access through a family account on each platform that allowed more control on access rather than relying on Google family link as some mitigation. The impact on young minds is yet to be seen from apps like tiktok which has such great algorithms that are amazing when properly used but reinforce negative messaging otherwise.
17. Limit the amount of time an ip address can access all of these platforms. Improve age criteria with these platforms my 9 year old set up a Facebook account.
18. "Make it mandatory for age restrictions to be in place & age assurance to used e.g a 10 year old should not have access to social media sites.
19. Encourage primary schools to engage it a mo mobile phone policy, if it becomes the norm no student will feel like they are missing out.
20. Parents have a big responsibility here too "
21. "More online training courses and awareness webinars could be made available to parents to Educate them on the dangers and on how to best protect their children, including how to set parental controls.
22. More tv and radio advertising on the age limits of different services (SnapChat, TikTok, etc) as most parents believe these apps to be for children and to be harmless. When so many classmates who are below the age limit of the services are using them, it is hard to be in the minority of restriction/refusing use.
23. More advertising on TV and radio signposting where parents can go to learn more."
24. More secure for parents to review and unlock before allowing child to unknowingly unlock the programmes.
25. Regulate the platforms taking responsibility for content and sharing.
26. I think video sharing platforms should be made more accountable for what is available on their platforms with large fines imposed on companies that allow unsuitable material on their platforms. They are not in my opinion doing enough. One on my children was exposed to sexual content on another child's phone in a schoolyard despite my child not being allowed a phone and age restrictions I have on devices at home. Blaming parents and making it their responsibility is extremely unfair. I also be in favour of the government legislating for this and banning phones in schools.
27. "Online short free course for parents, a standard, link sent by schools to all school age kids with basic instructions and info on the internet. I am a primary teacher, quite technologically literate. Despite all of this and my rules, my kids have still been frightened by ads for scary movies popping up or ads warning to report abuse if you see it etc. It is infuriating and makes me feel like a bad parent. The dream would be if you can develop an app for irish parents to filter everything basically!! Here is hoping !!
28. Thank you for this survey and all you do. "

29. Not sure
30. "Yes, there should be super strict rules and requirements for online platforms in terms of what is available to be viewed, depending on age.
31. I think the only reliable way to do this is to have mandatory, verifiable authentication in advance of being able to view content on, at least, the popular social media platforms.
32. I also think the current age of 13 should be increased to at least 15 for children to have their own accounts.
33. Children are, in many cases, not emotionally equipped to process much of the unmoderated content they see online and it can have a negative effect on their development and how they perceive real life.
34. Verification should be completed for younger children by a responsible parent/guardian before gaining access to a platform.
35. Would there be an opportunity to create a centralised identification platform that could use federation or a similar tech to log into sites once an initial verification is completed?"
36. Video sharing platforms must be held fully/accountable for the content they create/broadcast. All children are at risk, and additional needs children are even more vulnerable. My main experience is with YouTube Kids. I find the parental controls to be deliberately unhelpful. They will work on one device but can then be over-ridden on another device (Smart TV). The fact a child can simply input the answer to a multiplication problem to over-ride parental controls is a pathetic excuse at child-safety.
37. I think it's shocking that Ireland allows advertising addressed to children in ultraprocessed foods.
38. Make them take responsibility for what they putting up on their platforms. And who it's been aimed at. And should be no advertising to children, food, exercise body
39. There should be laws in place to protect children online
40. Every platform that is providing content to children should have, by law, parental control software built in. All content should have narrative descriptive keywords for parents to quickly read to help decide if the content is okay or not. I use Commonsense Media for a lot of my content information. We have devices in the house which have no parental controls and this causes problems. My children share my Audible and Kindle accounts and have free access to my entire libraries. It requires constant monitoring from me to make sure those devices don't suggest titles to my children that are unsuitable. I wish there were parental control options but there are none.
41. Kids accounts or age accounts should be colour coded, or have a very obvious symbol for parents to know the account is set up correctly, any child under 18 should have parental log ins, once a child has account it's hard to access them. All age accounts should have a similar theme or colour across the different app platforms
42. Moderate all content, such as ads, videos! As I found, there are lots of sexual videos on YouTube, that's why my children don't use YouTube, TikTok, Facebook and Snapchat. I find my children are not safe on those social websites. Thank you
43. Educating them, refresh inform them in school regularly! The reason is that the most of their time they are in school.
44. "Informing parents, including some info in SPHE for the kids.
45. None of the tech company CEOs' children have phones at a young age, which is telling "

46. Yes with parental consent
47. I think the commission needs to engage in an awareness and education campaign. As a parent, its difficult for me to explain in an age appropriate manner the dangers of inappropriate content and excessive use of online tools. If really appreciate some support in these areas.
48. I believe that the control and monitoring is up to the parents however all and any support from the Commission would greatly help to provide structures to online contente.
49. Not sure, but I do know that kids under 16 have multiple accounts, with multiple age's for different reasons...not ideal
50. I think Identification should be used before any account is allowed.to be set up and parental conformation
51. Accessibility for both posting and removing content should be considered.
52. I believe smart devices and social media accounts should be allowable only for those over the age of 18 or under with specific parental consent including identification documents uploaded by parents and that responsibility for the actions of minors accounts should be shared by the parents. I also believe there should be a way of parents acknowledging the use of the accounts, by way of contract or instructional videos which need to be watched before an account is created. Too many parents are clueless of what their children are at or have access to, placing children at huge risk from their own peers, other adults and themselves.
53. Yes platforms could automatically set standard parental controls on under age accounts and accounts where age I unverified. Stopping comments, requests. Messages from unknown people and stopping overage content featuring on their feed. These companies should also be held more responsible for what is on their platforms. When videos or fake accounts etc are reported very little is done about it. Often nothing at all! There should be harsher consequences for people who use the anonymity on these platforms to abuse others especially minors
54. "Influencers on sites like Instagram need to make it clearer when they are using a filter or advertising something. It should be displayed on the video.
55. There should also be more education around online and how what you see isn't real life. "
56. Age appropriate. No advertisements. Confirmation of child's age.Too many children have access to tik tok and posting videos
57. Ban Tiktok
58. Age appropriate content controlled by platform with heavy fines and controls in place by regulators, no advertising to under 18s, parental guidance on platforms and for devices used to access content.
59. "Parental education on danger !
60. Stricter requirements for platforms "
61. Parent education, online safety should form part of special needs overall supports.
62. "The you tube shorts are an absolute disgrace , I've searched everything to be able to block and there is absolutely no way! Yes you can block users of YouTube videos but you can't block the short videos. I think you tube is the worst app ever for children, they could be watching an innocent cartoon and half way through something totally inappropriate pops up. You tube seriously needs to be looked at!
63. The likes of tik tok, Instagram, Facebook under 16s SHOULD NOT BE ON THEM!! I think parents should block these apps for all their children under 16 on their phone. This is not just for the Commission for online safety to safeguard our children, parents need to be on their side too! "

64. Face recognition and age assurance are not an option as they would cause other undesirable effects. Gating is ok but parents need to be knowledgeable empowered and legally responsible for minors. Legislation on minimum age should be clear and enforced at home and in schools so that children do not feel that they are different or at a disadvantage if their parents are more concerned about their welfare and legality.
65. "Parental controls should be on and apps turned off where possible.
66. "
67. Online content should be policed better
68. Only allow it at certain times of the day.
69. There should be an age limit on using them at all
70. Kids shouldn't be able to see much of the content on these sites. Some of the content on kids utube is disturbing. I won't allow my 8 year old to watch anything on her own. Even for my 12 year old I am very concerned that he might go into content that is not suitable. I find it very hard to regulate this. It should be easier to block content based on the child's age.
71. Educate children in school starting at senior infants on how to use the online world and what to watch out for and how. The same way we educate children about crossing the road, strangers that approach them or any other danger in the offline world. If they know what to watch out for and how to behave they can always be safe.
72. Not sure, but anything to provide safety for the kids is valid.
73. The ads in some of the playstore games are not suitable for the age the game is suitable for.
74. Kids you tube is too babyish they won't use it. Then they go to their friends houses or the friends have phones and we have no control over what they see. They watch Mr beast in school. We are not parents anymore we are screen police and it is not healthy for anyone. Kids being exposed to all sorts eg erection ads on daytime TV or they try to stay up a little later over the summer but your sending them to bed for fear they will flick onto someone shopping for their next partner via the appearance of their body parts etc etc etc
75. Stop young kids being able to use tiktok etc. Primary school kids are on it because all their classmates are
76. More dialogue with parents and stricter controls
77. Parental controls assume once an individual is over 18 they don't need any filtering. There should be a way of having a setting for adults with additional needs still being able to have settings on their devices monitored by their parent/carer.
78. "The Commission should have a means to monitor and collate information from parents where issues are not being addressed by platforms. This would provide a means for industry monitoring and feedback to platforms on issues that need to be addressed (to be clear - this should not be a means for escalating issues).
79. In addition, every platform can attract bad actors and I find this survey amusing in respect to having different rules for different categories. The reality is that material that is inappropriate for children will show up in all categories sooner or later. In addition, bad actors will exploit any gaps in monitoring. Platforms need to ensure they are expending the same effort on abuse detection as they are on increasing revenue / viewing hours."
80. Children of all needs are drawn into these sites and have no control of what they will see next these sites easily drawn people down rabbit holes and can end up watching anything with the "up next" lime up is often very random
81. Prevention of harmful material being uploaded/viewed

82. I think that government needs to do more to protect children from harmful content and excessive advertising its frightening how addictive phones are and we don't fully understand the impact they are having on our children
83. Stop allowing people to friend people based on friend suggestions, I was horrified that people my son doesn't know could message him
84. "A parental guide to all ways and uses, safeguarding etc...."
85. I am fairly off with technology but it moves at the speed of light and it very hard to keep up with the changes.
86. With AI becoming more and more relevant we really need to up our game. We have no idea what is going on in the background.
87. It's very scary "
88. Some method on the phone that tells them - "why don't you take a break from your screen for awhile and go get some exercise /talk to someone " especially for boys
89. More regulation and education around social media platforms especially
90. While there's online security courses for parents available through the National Parents Council I think it should be included in the national school curriculum, the way that the stay safe programme is being taught. Now there might be an online security bit in that that I haven't come across yet, if so then the Stay Safe Programme needs to be highlighted more in schools.
91. "More how to for parents and kids
92. Updating information and options
93. There's always a new app or game - ways to keep up with latest trends "
94. I feel that by the time I learnt just how important this is, it was too late for me and my kids. I was of the attitude 'my kids are good and know what's appropriate' or 'they're only watching kids' stuff' (a bit of cartoons on YouTube). However, by the time they had progressed to using TikTok and other platforms it was much harder to then retroactively wrestle devices away from them and to install parental features, device time limits etc.
95. A smart phone ban for under 14's
96. "I selected "age gating" in the previous post because I wouldn't not like my children using AI / camera to "guess" my children's age.
97. I also wouldn't want to be uploading any of their personal data - like a passport to confirm their age.
98. My preferred method would be having a parent, add the child to a "family" account. And allowing parents to decide what age / category suits each individual child.
99. I have social media myself and the videos and posts I have come across on ALL social media is frightening. It doesn't take much to find -violence, gore, sexual, suicidal, hate, bullying and other inappropriate videos, none of which are limited to adults. I have reported numerous videos on Facebook and have had the generic "this has passed out safety standards" reply.
100. X (formerly twitter) has become inundated with horrendous videos of bullying in school, kids fighting and seriously hitting each other.
101. TikTok is full of dangerous "trends" which kids get hooked into watching as they're short clips. I see teens and pre teens who want to be "tiktok famous" and try re-enact these trends which can be very dangerous.
102. Snapchat is another app I dislike, kids able to send hateful photos and videos which disappear. Kids recording themselves doing awful things and saying awful things (bullying) thinking they can't be seen.

103. I think a lot of responsibility is on the parents too. Parents need to understand the dangerous around technology and allowing their children access to technology.
104. More courses in schools for parents would be great, safety nights, email reminders about child safety on the internet etc.
105. "
106. Parental information sessions. As a parent of children with additional needs, I have very little time to navigate the online world and keep up with all the new developments.
107. Phones themselves are causing huge issues for young people and parents on so many levels: they should be banned outright in primary schools and if possible restricted until child turns 15. See Jonathan Haidt's research.
108. "The platform my daughter uses is YouTube kids. She has been told to use this platform only. We have a rule she doesn't go onto YouTube without parental supervisions. My husband and I are not on Social Media so we are not familiar with Tiktok and Instagram. I feel there should be some regulation about the age of person before they can own a mobile device. No matter what controls can be put in place there will always be an individual that can work around this and still be able to access and share content. A national campaign on the recommendation of age before been given a mobile device. Pressure on parents is immense and also you don't want your child to feel left out or excluded.
109. I feel there is a complete lack of awareness on some parents part of the implications of giving your child a mobile device with access to everything "
110. Raising the awareness that regular & ncreased screen time damages mental health. A collective approach (parents & schools) to keep children off phones and screens would be very welcome.
111. Plenty of children are under age watching unsuitable content. Even snap chat needs the Commision for Online Safety to ensure age is real and not just entering a fake birthdate.
112. More control on what shared by the platforms is very important
113. Smart phones are as dangerous as cigarettes in my opinion and we need legislation to make it illegal for children under the age of 16 to own a smartphone.
114. Stricter monitoring of in school use of technology and more robust in school education on safe tech use. Tablets in my child's school were not monitored and children were able to freely download apps and access content that was not school not age appropriate.
115. "There should be no advertising whatsoever to minors online, not only things deemed generally inappropriate but also harmful to the individual or unhealthy, which varies widely from person to person. There is no way to fully monitor the damage so it should not be considered at all, it should all be banned for children.
116. There should be age verification on all content for minors that is age rated in any way above "all ages", & for those under 18 also parental consent. Anything inappropriate for minors should not be accessible to minors in any way at any time. All platforms such as TikTok, Instagram, SnapChat, etc. should require age verification & for those under 18 also parental consent."

117. I work in the safety org at Reddit. So maybe I am not the target demographic. Some of these questions were loaded in one direction or another. I would say that education is the most important thing here. All sites though have methods and tools in place to protect kids. If they don't then regulation should come from government. When people are educated on the tools available they will be more likely to pressure platforms into providing them. I also am not going to let my children have social media accounts until they are 16. This is not to say all content that is not age restricted is not suitable for children but that should be up to the parent to decide. By default all child accounts should be locked down as much as possible and the parent should be forced to removed restrictions as they desire.
118. "Hold the platforms more accountatable.
119. Enforce stronger age controls."
120. Ban these platforms from kids altogether it is the only way control access, my child has additional needs and he is well capable of getting work arounds to parental controls. So my attempts are futile. Snap chat is so risky as parents gave no visibility.
121. I am puzzled that this is about how to use such platforms rather than whether we should let children use them at all. I have answered that I don't implement filters on such platforms because my children don't have any access to such platforms and won't have as long as I can help it.
122. I feel that most of the time creators might not be true to the age restrictions of their content. This could be because they want to drive as much viewings as possible as they will reap benefits from it. That being said there is no true classification of content that will actually stick to the age profile. I've seen this with Youtube Kids where I doubt some of the content has been verified before placed in the platform, as the creators have to put their own classification. So I think the platforms have a big responsibility to accurately classify the content, as the creators do. My suggestion is to use a moderator that can verify the classification and change it accordingly with specific rules... or you can use generative AI to analyse the content and verify that same classification, having some human help on those cases where there can be a doubt.
123. There should be a legal age limit for certain usage and time aloud/limit on each platform.
124. Yes these platforms should provide free internet safety talks in schools
125. In an ideal world, we should not be handing out a super computer to children under 16...at least their brains might be more developed by then
126. Enhance awareness about parental controls AND on how to use them. Share the obstructive, provide some basic training videos share them online make it as easy as possible. I've had problems with youtube kids parental controls and set up so it would be great to have support
127. This needs to be a priority for all company's providing video sharing platforms. So far they have got away with too much and need to be held accountable. They need to enforce stricter age limits on material & any inappropriate content needs to be removed immediately. Twitter has got rid of most of its monitoring staff for this and this is not acceptable. Children are being exposed unnecessarily to inappropriate content and this is going to have a huge impact on them developing into sensible adults.
128. Education on safe usage. Showing stats on how long they spent on it and categories of usage they spent their time on
129. Video Sharing Platforms need to be held more accountable for their content and who it is aimed at.
130. Encourage children to limit phone usage.

131. Limitations to the amount of time they can spend on them.
132. Yes
133. I think if age related controls could be implemented , many of my kids friends had access to tiktok , snap chat at young ages , as girls can easily look older and they all entered false dates of birth. I think a lot of harmful toxic media content , should be age 15 and above and enforcement should be tighter, as despites having parental controls on apps , on qustodian which have blocked my older child from being sent porn . other kids have shown her the images / content on their phones . So even though I am trying to limit / control these on my daughters devices . I have no control over her friends devices and what they show her.
134. Education for children to help with judgement as no matter what control are in place you cannot assume everything harmful will be stopped while you can't fully control what they see on peer devices. Training for parents on parent controls and how to work with their children to monitor usage, discuss content and build good behaviours and judgement around social media given . I think the scope should also bring in AI generated material as it is being incorporated into search engines and productivity products such as MS office tools.
135. A child can enter porn and is exposed to all types on this
136. Unfortunately in todays Irish society children with additional needs are often targeted by bullies using the platforms mentioned above. More needs to be done by the online platforms themselves to prevent this happening. Social media and online platforms need stricter monitoring and controls in place to prevent them from being used by others in this way to cause harm.
137. Clear simple reporting procedures for inappropriate content should be in place accompanied by clear guidance on risks and appropriate controls
138. Ban Social media for under 18s
139. I think that children under the age of 18 shouldn't have access to these platforms and the person using such platforms should be required to submit their passport and verify their account with their finger print or facial recognition.
140. It would be helpful if controls were in place by default. We control the content our kids watch, but even on Netflix content rated U can be inappropriate.
141. The best advice I have come across in relation to this is to watch the online content together with your child as opposed to just throwing them a screen to keep them quiet. Often, there is inappropriate content on seemingly harmless videos such as make up and beauty etc. so even with parental controls in place it is very difficult to keep on top of what your child is viewing online unless you watch it together for a set time i.e. 1 hour a day.
142. My main concern is Snapchat as the messages disappear. My son has only expressed interest in this platform joining secondary school. thank goodness the school have a good policy to mobile use.
143. Bring in legislation banning children from using video sharing platforms and until age where these platforms are least harmful to children
144. I would like to see greater punishments given to content providers who blatantly break the rules. Fines are irrelevant due to the huge incomes they create. A break in the service being provided would offer a greater deterrent.
145. More should be done to control what's posted, even with parental controls on on channels such as YouTube, I have seen videos where people found a way around these controls and included inappropriate content. For example, a kids video showing someone playing minecraft and suddenly the person videoed the phone in their hand and on that phone was an adult video playing. So that happened half way through that minecraft video.

146. I would like to attend training on how to set these parental controls and monitor my children's online behaviour
147. More ability for parents to limit content based on their own expertise e.g. YouTube is the only sharing site that my 10 year old uses but is not allowed their own YT account based on their age(by YT). However because she uses one of our accounts it's hard to control the advertising though I'm monitoring what she watches.
148. Actually after completing the survey, I realised Parents like myself could do with an information session to educate us on how best to keep our young people safe online.
149. Regular workshops with professionals, journalists, psychologists, high follow influencers. This way children can hear more aspects from different angles. This non-educational rather discussion based sittings suggest common sense choice in children's behaviour.
150. Make the platforms accountable for the content they show the same as traditional media
151. Yea the companies should contribute to child mental health services and child physiotherapist as no matter how much we try, we are losing an entire generation to social media
152. Block advertisement entirely.
153. Parental controls, ability to turn off advertisements for children especially those with sensory difficulties.
154. I have a 14 year old who self verified herself as a 22 year old and while accessing chat channels was exploited online. Despite being a minor and legally not able to provide digital consent in Ireland, internationally none of the platforms I contacted - Reddit, Discord, Twitter or TicTok accepted any responsibility for what happened to her. In their view, none of their 'policies had been broken' due to her self verification. They wouldn't even take down images despite my pleas. It is very hard to balance privacy and freedom of speech with child exploitation, particularly in private chat rooms. At least TicTok have some measures e.g. you can't share images privately. If your child is willing to accept the platforms parental control boundaries, then you have some chance of 'controlling' what they see. But if you have a digitally literate, curious child - you have no chance! The platforms need to put more safety measures in place for adult content - particularly porn which is increasingly violent and denigrating to females (and that's not me being an old school prude). School's SPHE programmes could also do more to counter this online portrayal of sex which is unhelpful for both males and females. Teenagers are learning unhealthy images which then create unhelpful expectations of sex e.g that is ok to choke or be choked. We really are sleep walking into a societal time bomb and it is not surprising that youth mental issues are on the rise.
155. "Expressly forbid devices capable of accessing Internet in primary schools other than school devices.
156. Funding for annual training in cyber safety for all teachers and pupils from age 9 / 3rd class.
157. Discourage use of phones in 2ndary schools - eg to access curriculum."
158. A module for kids in school and an online module for their parents

159. "There are monitoring subscriptions available however these are not fully usable on iOS due to security. There should be a legitimate option to bypass built in security measures

on iOS so that these third-party subscription monitoring services can allow parents to fully monitor child's internet activity on their iOS device.

160. Also, I feel the 'disappearing messages' format of Snapchat is inherently dangerous and ripe for abuse by bad actors. I believe these chats should be backed up on a Transcript that can be viewed by a parent. "
161. "Recommendations around mobile phones in primary schools (i.e. smart phones not to be used by under 13s)
162. Parents need to take more responsibility for their children's online presence, become more familiar with parental controls etc.
163. Supports should be inclusive for all as standard. "
164. No
165. Moderation of videos should be much improved but also kids should be learning in school and at home about how what you see on these platforms is not real life, it's filtered, edited, advertising, promotion of a person etc.. weekly open discussions in schools in every class at an age appropriate level
166. More education needed in school regarding the dangers of online content. Children find ways of bypassing all the safety features available to access what they want.
167. If a child is uploading a 2nd party should approve before it can be uploaded
168. I think children should be treated equally regardless of additional needs.
169. Its parents responsibility to filter and control what the kids watch nowadays, do our best.
170. Safety of kids first. Default settings should do that. Should not be relying on parental knowledge
171. I don't think fining these organisations works because they are generating such huge amounts of money, I think there needs to be a more effective way to make them responsible for the content.
172. Disappearing messages are a big concern!
173. I believe there should be a national policy for disallowing electronic devices in schools similar to the scheme which was introduced in Co. Waterford lately.
174. It would be very useful if there was a video or other online training for parents on regulating their child's online usage. How-to videos etc. on setting up these parental controls would be helpful. Also coordination of these controls among friend groups would be ideal as my daughter regularly says that she is the only child in her class with online controls, app restrictions etc. I appreciate this would be hard to do.
175. More detailed parental controls which respond to issues that occur on platforms, clear advice to parents on what age platforms are designed for e.g. YouTube, Instagram and Tiktok are designed for 13+, yet 8 year olds have their own accounts. More targeted info campaigns regarding online scams on these platforms for those with additional educational needs online as they are exceptionally vulnerable. As aside but still relevant: Requirement on shops selling devices to help parents set up the device correctly, activate controls etc. and also looking at the pre-installed Apps which are on devices used by children. The influence these platforms can have on younger children buying products online is also an issue. Thank you.
176. "We have a child with additional needs.
177. Better educate parents"
178. Run practical courses for parents - get them to bring their devices into an accessible class and show / demonstrate how to use parental controls. This should be a hand on / practical class.

179. Advise setting PIN codes on adult's profiles. Remind people that you can block certain programs in a child's profile on streaming providers. Advise that kids YouTube is never 100% safe due to the way content creators try to get around restrictions. TikTok should be restricted to 12 and upwards, due to the dangerous "challenges" that often appear.
180. Have tighter restrictions on who is uploading videos and what content is in said videos
181. They could regularly keep a check on the age group that are using these platforms not and regulate all platforms so that the age group cannot go into content that they are not supposed to look into
182. Accessibility is an issue. Kids are more tech savvy than their parents, so safeguards need to be there to assist the parents in safe management of device use & content access.
183. The rise of deep fakes in light of rapid AI developments and the amount of fake news is a concern, especially as I see young people getting most of their information exclusively from online sources. How can we help them differentiate what is real and what is not?
184. When flagging bullying, follow up should be enabled which includes a consult with a therapist, blocking bullies from communicating, informing bullies parents
185. Age limit and parental consent
186. Ensure they can not have accounts under the digital age of consent. Ensure social media providers apply parental control on adding friends on younger children's accounts. Parents vet friend requests of their child so they know who they are talking to online.
187. "Yes. Make educational content (video, presentation, etc.) on the topic of Internet safety and send out this content to parents so that they can discuss it with their children. Or organize a lesson at school (using prepared educational content) on the safe use of the Internet for children. You can also use these two methods at the same time.
188. Thank you for taking care of our children."
189. Dangerous content like abuse videos, bullying etc should be banned and taken down. All content should be vetted before upload to TikTok etc. There is insufficient barriers in place for children even on Kids YouTube they can be exposed to inappropriate content.
190. "Point to reliable and safe sources for key information related to content viewed e.g. HSE.
191. Prompt the child to talk to a safe adult if they are affected by anything they viewed that confused or disturbed them.
192. Clearly state that the content is for people aged over YY and that if the person is younger, the content may be quite confusing or upsetting. "
193. Provide training and information on social media and having an online presence. There is no getting away from social media so why not arm them with the tools and knowledge to use the technology responsibly and get the benefits of it
194. Run workshops in schools for staff and parents/guardians on online safety.
195. not sure
196. We should follow the UK model of adult content only being made available if you specifically request it from your ISP or mobile operator. This is also only available to those over 18.
197. If a young person with additional needs post an unsuitable footage it must be taken down no matter what age they are
198. n/a
199. Parents should enter a pin for any unsuitable content for kids age that's inappropriate

200. Teach them how to use it in school. Teach them like a subject and monitor how they understand it in school.
201. More education aimed at parents.
202. some stakeholder other than the parent / Child needs to limit the amount of "on-line" time children can spend on a device daily. Children don't have the self control to manage this & most spend hugely excessive time on-line. I feel Most parents aren't technology savvy enough to managed (unless they work in the IT field). I feel the result is having a damaging & negative effect particularly on 12 to 16 year olds daily lives.
203. Better monitoring of spam/advertising accounts
204. "I think you have covered it all, however as my daughter goes into 3rd class, I am concerned by the level of bullying that is happening online in chat apps etc. and I believe that this needs to be considered as seriously as the online access to sharing platforms.
205. Many thanks "
206. They could try and enforce a way that the owners of these platforms should require confirmation from parent or responsible adult to prove the child is allowed to use them
207. Enhanced parental controls, different age recommendations
208. Introducing an online platform where users can report safety issues not addressed or incorrectly addressed by the video sharing platforms, for further analysis and action against the platform, if needed.
209. Education. Online literacy, safety and supervised practical experience should be incorporated into all levels of the school curriculum (plus homework exercises involving parental participation). Not just using devices to complete other parts of the curriculum (e.g. maths, reading) but a dedicated 'Digital Life Skills' subject to compliment traditional Home Economics.
210. Access to bad content should be locked and only made accessible by a department person who can verify the person is an Adult and should be done monthly in case of a child breaking into an Adults site.
211. I've no clue on technology. Very basic. & shares me children know more
212. ensure schools technology education is focused on staying safe online, thinking critically and evaluating digital info, social media caution - these to me are more important than using digital info as I think all children now are exposed to technology and being aware they need to be cautious and limit its impact are more important than digital skills for this generation.
213. This is a priority for any parent and social media and digital platforms are causing a massive negative impact on the lives of children and teenagers. I believe this topic should be high on the agenda!
214. I have been using parental controls and selecting content on age rating but still I find too many instances during programmes where content does not correspond to the rating given or content explained at the beginning of the programme. Online platforms don't do enough to identify and filter material based on rating. Children still get exposed to inappropriate material during programmes which have been targeted towards that audience. For example how on earth gun violence, suicide, gory images and nudity with sexual content ok to watch for a 13+ or even 15+child? Still many programmes aimed at 13+ and 15+ show all this in the programmes. When it comes to using copied images and soundtracks, online platforms identify them with their AI algorithms because it affects their revenue but there's blatant lack of responsibility and sheer inaction on their part which is leading to mental and psychological difficulties in our present and future generations. These online giants must be made responsible to do more towards ensuring that young minds get healthy entertainment.

215. Provide a little more assistance
216. If possible educate children about online etiquette.
217. Common parental controls across all platforms, setting parental controls on one device or platform eg google, does not populate it across all platforms and is impossible to gauge how safe any platform is
218. Just a good training for parents regarding parental control
219. "Ban Tik Tok, Snapchat & instagram. It is destroying our children's lives. The content is ridiculous , gives them access to everything everyone and anything,
220. children believe the content is true, & older people are contacting children offering "videos"
221. Of a sexual nature.
222. This country will have a very serious problem in a few years if there isn't something done now to protect our children. All phones with should be banned till they are at least 16, or just have a phone that can ring & text. This is a crisis situation but nobody seems to notice"
223. It's too easy to give a wrong date of birth. Should give ID for Snapchat and gaming
224. I found even with Parental Controls we have come across adult content especially YouTube and Tiktok
225. Stricter rules around children using sites
226. Work with primary schools to discourage students bringing phones into schools
227. More training for parents in how to restrict access to content for kids
228. Normally the parent should be responsible to limit screen time as I do with my child with additional needs, and should be very attentive to what they watch.
229. Not that I am aware of.
230. Completely disagree with question 12(Did not answer). The parent should upload their own documents or verify a child's age. Parents must also hold a responsibility. Why would we want to upload pictures & date of births of our kids to online streaming platforms. Raises a red flag for me.
231. Should have parental consent to access these platforms and that is for all children, not just those with additional needs!
232. Help in educating parents on how to have more control over what they are looking at and educating kids more about online safety and that most of what they are looking at is not real life!
233. Force online platforms to moderate the content published on their respective platforms in line with the age appropriate content guidelines laid out.
234. I think it's too easy for young children to get onto certain apps all they have to do is lie about their age. I have also found that on some children's apps/games that their are people on their messaging inappropriate things. I'm lucky my child told me but not every parent will know the things being said to their kids. He was on Among Us game which I thought was safe enough obviously it was deleted straight away.
235. I'm not sure. I don't believe young kids should have access to these platforms at all.

236. Educational content to warn them of dangers needed
237. More regulation needed from the top down. Parental controls should be more accessible. Sometimes they're almost hidden within the app
238. Nothing comes to mind
239. Ensure schools enforce rules and discuss them with parents and children

240. Enable voice over messages sonuser is aware of content age the content is aimed towards
241. No
242. Work with Media Literacy Ireland
243. keep photo thumbnails when search results come up as alot of special needs relate to front pictures of the video or song they are looking for. Also it would be good to have an option that says click here to "skip add" with an arrow as sometimes my daughter can't work out when to press to skip the add as some you have to watch the add to the end and some you can skip after 20 seconds
244. Provide child-friendly videos that educate children on the value and dangers of online usage. Children need to see and hear a voice other than the parents. Perhaps have small infomercials before or after the news or children's programming on television.
245. Not sure
246. Make sure reported content is handled promptly and correctly, giving feedback to the person who reported these.
247. "The video platforms have to be positive and safe for any online users.
248. Children videos must have safety control on Ads and misleading information about promising things that can't happen in the real world.
249. "
250. Passport should be used to know the age of the child and only information for the child should be send to the mobile.
251. Yes - provide free advertising for childline and associated children's charities and helplines
252. Ban smartphones from schools
253. More laws should be brought in to protect innocent children. Their young minds are unable to process so much content. This can be harmful.
254. "Education and informative ads about parental controls and inappropriate content should be placed into those online places where they can be seen and targeted i.e. if you've watched 1 hour straight on youtube kids or roblox an advert should interrupt asking them to show it to their adult before proceeding.....
255. also short tv adverts on telly eg at news time or during coronation street, as lots of grandparents have laptops and tablets and allow kids to use them without understanding the risks. They certainly wouldn't have the tech savvy to start setting up different accounts and different controls etc"
256. Online Parent education
257. Maybe courses for parents about how to use Parental Controls
258. While I appreciate the intent of age checking I do not want to provide id and birthdays to strangers online. I much prefer other controls. Unfortunately it is mostly the responsibility of the parent to ensure kids aren't able to access inappropriate content, and information on how to do that should be more widely available.
259. Age assurance should be required for social media apps like Facebook, Twitter and Instagram. Greater education in schools needs to take place and greater education for parents to teach their kids about online behaviour. Sellers of digital devices should be mandated to provide parental controls on all devices.
260. Videos etc for Tiktok, YouTube that starts with appropriate content with very inappropriate content within to fool the parent controls. More moderating please
261. Be more vigilant in ensuring that companies remove harmful content

262. Ban phones from schools, require phone manufactures to provide age appropriate operating systems for mobile devices
263. Age verification using government identification for all Social Media and messaging apps with a strict user age of 14 years
264. "
265. We don't let our child on tik. Tok YouTube YouTube kids our Instagram. Full of total junk and simply not safe to not be sitting there with them.
266. Suggested age gating as tbh no company will monitor passport upload etc so in reality can't see it working. Doesn't mean I think it's right.
267. <https://culturereframed.org/> should be shared with everyone. Useful clear guidance to stay informed
268. Good luck
269. "
270. I think both the content creators and platform should be responsible and accountable for the content created and shared. There should be strict laws in place to enforce the rules.
271. Vigorously promote the widespread implementation of the Greystones initiative in primary schools. No child should have unlimited access to the internet / social media etc. etc. etc.
272. I'm not sure
273. Ban them !
274. Only allow them to be available for 18+ and putting the responsibility on electronic device companies, media platforms and those who make the posts etc
275. Produce one trustworthy document that parents can access easily, to explain how to monitor content and set limitations on each of the main platforms. Allow comments on same so parents can provide each other with additional information.
276. There needs to be education given to all parents and it should be mandatory. The school I am in provided Internet safety talks. It was brilliant but I know from attending that the parents whose children have phones were not there. Children in my child's class are using certain apps which are dangerous and not suitable for any child
277. All apps should have a "Grown Up" mode that an adult has to enable so childrendont get exposed to content inappropriate for their age
278. I think they should all be banned imo.
279. Better monitoring of online bullying
280. Smart phones should have an age limit 13+
281. Additional advisory warning. Allow and act on feedback from parents.
282. Making sure there are no gaps or missed content that are not age appropriate
283. The age limit should be higher and only accessible through more strict age verification proress
284. Legislation needs to be put in place to inforce policy rather than it being discretionary
285. It is my opinion that no child under the age of 16 should have access to social media. There is absolutely no evidence to support it being beneficial in any way whatsoever. Social media accounts should be linked to official ID documents e.g passport or driver's licence. You can only set up a social media account by providing verification of either document. Children under 16 should not be allowed to have social media accounts.
286. "1. The age of digital consent in Ireland should be increased to 16

287. 2. Classes for parents as none of these things are failsafe and parents need to be more vigilant/discuss with their child more what they might see etc.
288. 3. More on the dangers... Particularly around online bullying leading to suicide. Online eating websites leading to body dysmorphia and disordered eating
289. Consent and tell not to video share, even with trusted people.
290. Educate parents in simple lay man's terms through the schools network do parents realise we are all in the same boat trying to protect our children
291. I wish all phones were banned for use in secondary schools. They have no purpose on school grounds. My daughter is asked to use for Google classroom and says they all just go on snap chat. I wish smartphones were illegal for under 18's.
292. they should be able to restrict these platforms to age appropriate. So if you are 6 that you don't have tiktok. Snapchat should be banned its a purely bullying platform. they can group chat on whatsapp if required. you tube should be age controlled.
293. As far as I understand inappropriate videos at the moment have to be reported by many viewers before they are taken down; I've age restriction on my kids YouTube yet I don't seem to have access to what they've watched; on Netflix kids account seems to allow stuff that's not age appropriate; Disney channel has had wrong age ratings for movies - weren't actually kids movies and showed "U"; movies like Home Alone and many more older movies have violence in that I find inappropriate for young audience or even for myself; there used to be lots of fake Peppa Pig videos on Kids YouTube, and many videos with hurtful underlying/built in/hidden messages - I'd love to hear that there's a way to eliminate those videos; at the moment, when I want to block a channel on kids YouTube, I need to start the video, to get "block the channel" option= there should be an option to block these channels without watching any videos, and to block channels with specific area like video games etc that don't suit some families; I've YouTube premium for the whole family= no ads, therefore I can't comment much about ads, I just know some Sky apps like Ninja kids show ads, they seem ok so far. Myself and my kids have stopped watching YouTube many times. Unfortunately on TV there is no option to delete Kids YouTube app. I understand sometimes it's useful, even school recommends some alphabet stuff. Thank you for caring.
294. No comment as I'm not parent of child with additional needs
295. maybe a way to link accounts privately so any harmful content would be flagged
296. Yes Tiktok is totally unregulated and I would argue does not have a sufficient process of verification of account creation- I currently have an open case with data protection with them whereby they enabled a fake account to be created using another person's data to purport to be a young girl. I am frustrated with the fact that providers self regulate and parents are oblivious to the fact that children are essentially using self censorship. It is a slow process to get any accountability and regulation is too slow and unfortunately not implemented including GDPR- parents need to be educated as to the dangers of children and adults giving away their data unintentionally or data theft including personal identification data I also know a friend whose facebook account was hacked purporting to be another person and they also cannot get any accountability - it is very hard to get any support. I also object to the wording of one of the questions whereby the answer may be skewed in favour of platforms indicating data is tailored to individuals - there should be more options or a free comment text answer for this as it only allows answer stating they are not aware or are aware as such agree with the comment.
297. Promote no-phone policies in schools
298. Visits to schools to provide info sessions on these platforms/ simple visuals or video clips to explain

299. "Provide classes to teach parents exactly how to put on parental controls. would be useful.
300. Getting a law to ban phones for primary school children and ensure they are not used during school hours in secondary school. They are harmful, stricter rules can be applied but it needs to be same rule for all"
301. "1) our culture, and state policies should affirm that it is the parents' right and responsibility to screen content according to their wishes, for their own children - and options (technological and social) should be promoted to facilitate that parental choice;
302. 2) it should not be the first option - by state and society - to mandate that everybody else and every sphere of online activity has to bend towards being a functional babysitter for everyone's children - and turn into a virtual panopticon against adults and children both."
303. I'm very pro the banning of mobile phones for primary school age kids. If we all do it then nobody is left out. It is just how it is. Really like what they did in south dublin (I believe). Can we make it national?
304. Advocacy issues
305. "Best is for children to not have access to tablets, phone etc ... and banned social media platform (tik tok, insta....) as they don't bring any values to education or development.
306. "
307. Creation of an online platform aimed specifically at primary school age children where the parent must approve/deny posting of pics, posts, etc on a live basis. That way if inappropriate content is being produced, an adult is responsible. The platform should also provide mediation of some sort for any disputes which arise where the parents/guardians can discuss the post.
308. Rather than trying to police the use, which children can figure out how to bypass, inform children about the motives behind the tech industry, that using video sharing platforms is free because THEY, the children, the users, are the product. Some platforms can be very creative with video editing features, children engaging with these features might help encourage children to be more active in their use (creating) rather than passively scrolling.
309. Have the ability to link these platforms to a parents mobile/tablet so that the parents at all times have the power to oversee exactly what the children are watching and getting involved in
310. The commission should advocate strongly for the government to introduce strict controls on content for platforms. I think they should be regulated in similar ways to broadcast media
311. Phones are the main issue for me, kids have access so easily. We haven't given our 12 year old daughter tik tok yet but she still gets sent stuff from the platform from her friends. Also, use app which we pay for to control phones but the apps will cover something and often cause issues with the phones.Very frustrating.
312. More information needed on parental controls for different gadgets such as laptops vs phones etc and how to link these as some apps may only work in one setting.
313. More education to parents on the true harm of these platforms to help them steer their children away from them for as long as possible and co-use them thereafter. Even adults struggle to monitor their usage habits - underdeveloped child brains haven't a hope against the algorithms.
314. "1. They could make it compulsory for parents to attend a 'how to keep your kids safe online' course.

315. 2. They could fight to make it illegal for children under a certain age to use a smart device.
316. 3. They could fight for child friendly phones be invented."
317. I wish none of it existed for my children. Business benefits but no person benefits for kids
318. Work harder to keep children away from social media until they are in secondary school. Provide more training and resources for teachers and parents. Have more advertisements related to the potential harmful affects of all of above.
319. Create an educational environment where those platforms are eradicated as they have no place in the classroom
320. "Fines for the platforms who are not enforcing age restrictions.
321. Platforms should be monitored by external bodies which are government funded and should have the ability to issue large fines where harmful content has not been handled in effective or appropriate ways."
322. Perhaps in-school talks to kids making them aware of the dangers of online platforms, misinformation and real life examples of harm/problems caused by regular use of platforms. Include age appropriate talks on the addictive nature of video games, pornography and the truth behind the pornography industry and the harm viewing this, violent games and over sharing by children (photos and videos used to ridicule etc) can cause using real life examples.
323. "Fine the platforms if content isn't taken down after being reported.
324. "
325. It should be law that kids under 16 have no data/internet on their mobile phones. Smart TVs are also an issue going forward as they all have streaming platforms.
326. A default Automatic 'time out' of social media platforms after 1-2 hours, unless settings are changed
327. "if we are going to take on line safety seriously, we understand the risks, the devices on which underage children can access such content should be better regulated... if you have to be over 18 for social media, there needs to be much more robust regulation and penalty where underage use is identified and parents need to play their part in this.
328. Our children need to be given the opportunity to be innocent, and also space and time to develop their own views, not force fed constantly by on line vultures.
329. Technological advances were intended for the adult working world to provide efficiencies, NOT to e a toxin for our youth. "
330. Stricter controls on targeted advertising
331. Bring in new Laws so the makers of harmful content are prosecuted.
332. Educate children and adults. Ongoing information at school in the community.
333. I think it needs to be done as soon as it is possible for children
334. Have direct connection/ regular exchange with the platforms to ensure their collaboration on the matters. Most of them are very interested to get this right for the right audience
335. There should be a state platform, an RTE version of YouTube. No likes or comments.
336. We shouldnt allow phones in the primary school .
337. Support the parents by providing workshops to explore issues
338. Why can't we do what France did and put a legal age limit of 16 on mobile phones/tablets. So kids can't own one legally u till then. Then I think there should be

consequences for publishers of inappropriate content not labelled/restricted from children in these platforms

- 339. So scared of what internet can do to my little ones....
- 340. Ensure that content can be taken down when it causes hurt to someone.
- 341. I believe protection for children from harmful content begins with how the smart device is set up by the parent to tailor the restrictions to their age. Providing how to guides to parents would be the best support for parents.
- 342. stop watching you tube, tik tok Instagram children under 15 of age

Children's comments responses to question 19

1. Some children and young people need extra support for reading, writing, hearing difficulties, difficulty seeing or other types of difficulties. If you need extra help for deciding what videos to ...
2. Should be able to email or notify the companies if you have any of these difficulties
3. No comment.
4. "No
5. "
6. unsure
7. no
8. Add subtitles and sign language. Audio Description of video for blind people
9. Yes, please put more clear and simplified rules for content sharing, links and how paid advertisements mid videos work.
10. Toodloo
11. I don't need extra help
12. Auudio description so they can hear what it's about not just read it.
13. idk what ur on my moms making me do this
14. Sign language writing it out what they are saying in brades
15. I'd like to press a button for it to be read out if I can't read it
16. Yes
17. I think if you have trouble reading that you would not be on those apps. But there are some supports for that. Also the apps are catering for the majority.
18. Im not sure
19. I think that you should be able to have a filter that shows someone translating videos into sign language for people with hearing difficulties.
20. Let me watch what I want
21. Small description to let you know the content
22. Subtitles
23. When making the video put the oldest age that can watch it.
24. All videos should be checked by a safety person from the platform before they can be approved to be uploaded online
25. Have a section where you can highlight need for extra support, where you can choose what you require eg. subtitles maybe.
26. Easier way to remove undesirable content

27. Maybe have boxes come up on screen before videos to suggest getting parental advice before playing the video, like a second opinion
28. yes
29. Sometime language is not proper according to age group so need improvement in this case
30. A setting to say edit what they think you to watch or edit the content field with tags or ban certain tags or disable content feed as an option and YouTube should add back the dislike counter so that I can see how good the videos are and also the ability to rate ads with a 5 star system.
31. Additional parental controls. Tickbox to say the user has Additional needs
32. I would only like to watch video games because video games are just made up games on PlayStation, Nintendo or Xbox.
33. Age appropriate content; that everything is checked before it's posted on YouTube; clear and longer description on YouTube kids videos.
34. The video should be verified by a teacher and have a symbol on the video so that they know it will actually be able to help them
35. Have a section for easy access with supports for anyone with extra needs so they don't have to scroll through all the videos
36. I think that there should be automatic subtitles to be turned on for deaf people or people with worse hearing

Keeping our children safe and informed when watching online videos

595

Responses

22:37

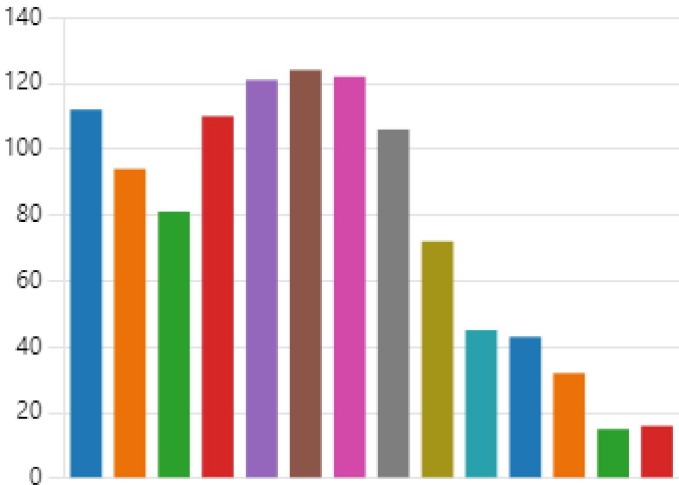
Average time to complete

Closed

Status

1. It would really help us if you could tell us how old your children are, please tick all that apply.

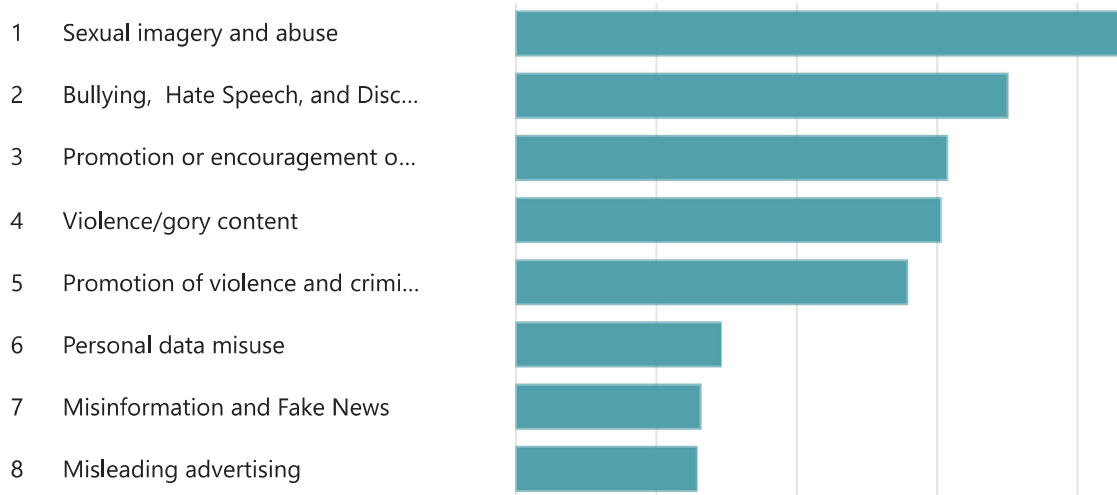
5yrs	112
6yrs	94
7yrs	81
8yrs	110
9yrs	121
10yrs	124
11yrs	122
12yrs	106
13yrs	72
14yrs	45
15yrs	43
16yrs	32
17yrs	15
18yrs	16



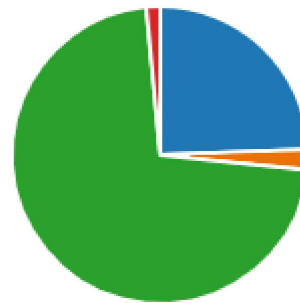
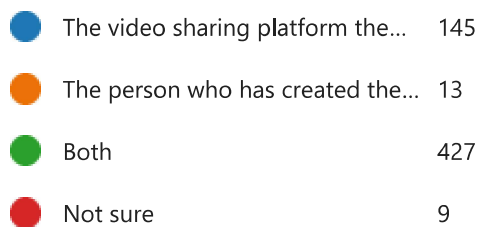
2. When our children are viewing video content online, we would hope that they are finding the content enjoyable and educational, but sometimes their experiences may have a negative impact on them.

In your opinion, what types of online content would concern you the most?

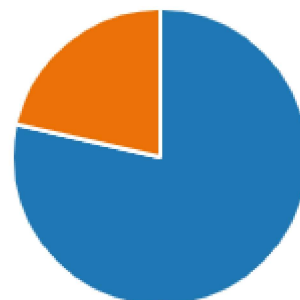
Please drag the options to indicate (in your opinion) the most concerning to the least concerning



3. Who do you think should be responsible for regulating content on videos shared online?

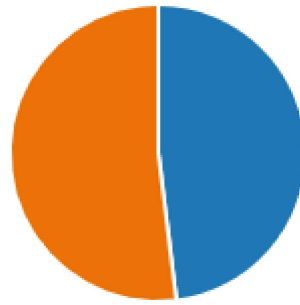


4. Did you know you can report harmful content on video sharing platforms?



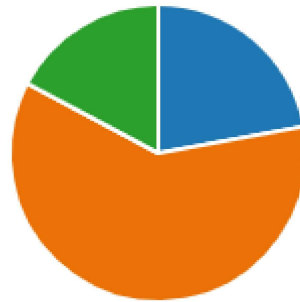
5. If yes, have you ever reported anything?

● Yes	223
● No	241



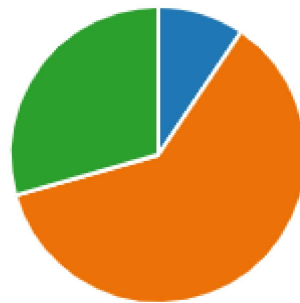
6. Were you told of the outcome?

● Yes	72
● No	197
● Other	56



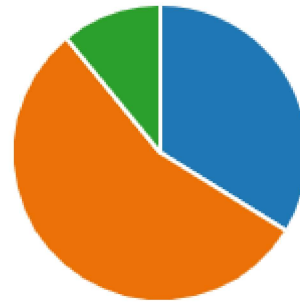
7. Were you happy with the outcome?

● Yes	28
● No	181
● Other	87



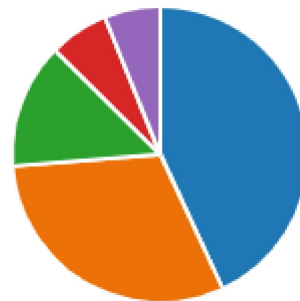
8. How familiar are you with content rating systems used in movies, television, streaming services, and video games?

● Very Familiar	200
● Somewhat familiar	327
● Not Familiar	65



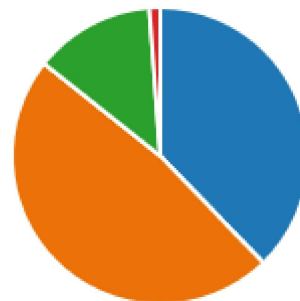
9. What types of content rating do you think are most helpful in deciding if a content is okay to watch for your children? (Please select all that apply)

● Age-based ratings (e.g., G, PG, 1...	319
● Descriptions of specific content ...	227
● Parental guidance information	101
● Viewer reviews and comments	48
● Other	45

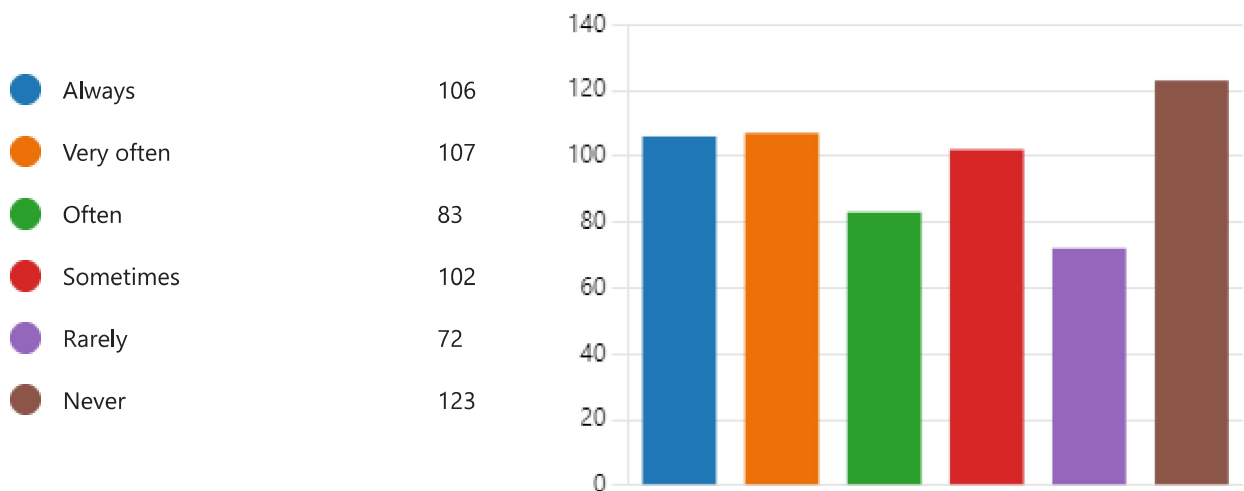


10. Are you aware that you can use content rating information for selecting content on different platforms (Such as YouTube, TikTok, Video Games, Instagram)?

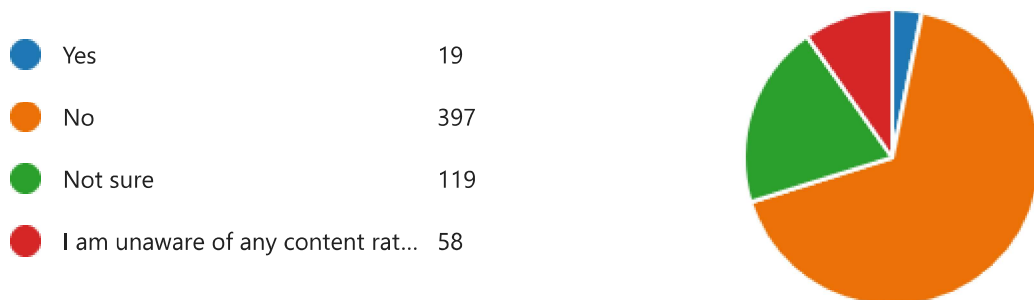
● Yes	224
● No	284
● Maybe	78
● Other	7



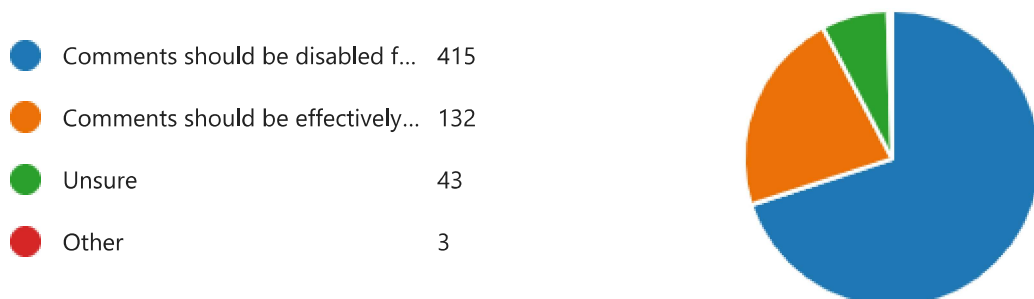
11. How frequently do you use content rating information when selecting videos or monitoring your children's viewing on Video Sharing Platform Services?



12. Do you believe that video sharing platforms provide enough information about their content to allow users to make informed decisions before watching them?



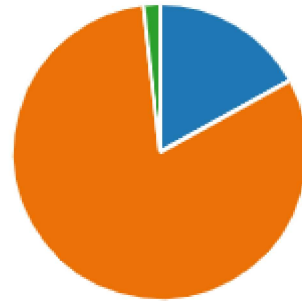
13. How do you feel about comments linked to videos intended for children? Please select the option that best represents your viewpoint:



14. How do you feel about sponsored content (advertisements or promotions) that is integrated into videos aimed at children on platforms like TikTok, Instagram, and YouTube?

Please select the option that best represents your viewpoint:

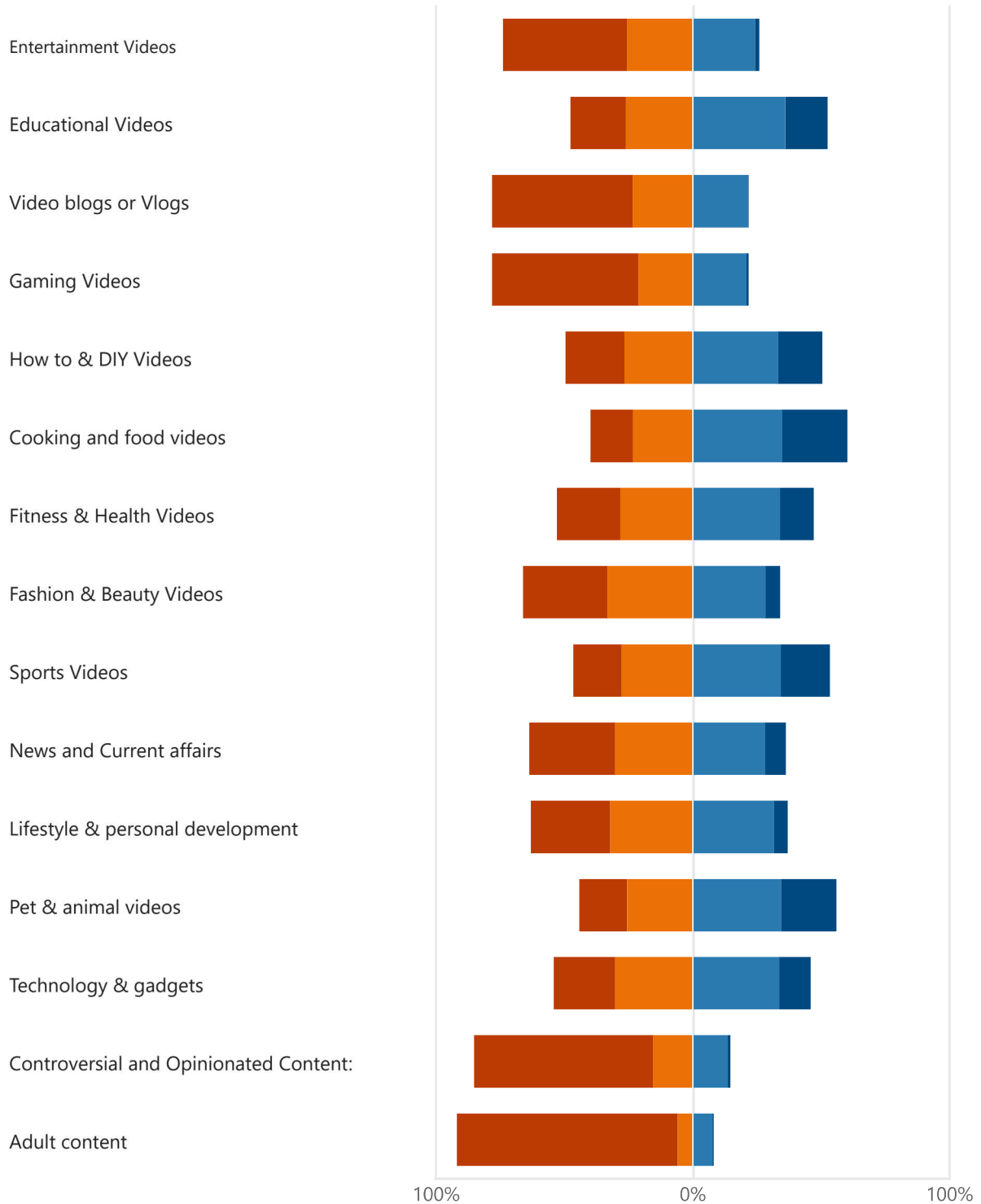
- I believe that sponsored content... 101
- I believe that sponsored content... 482
- Unsure 11
- Other 0



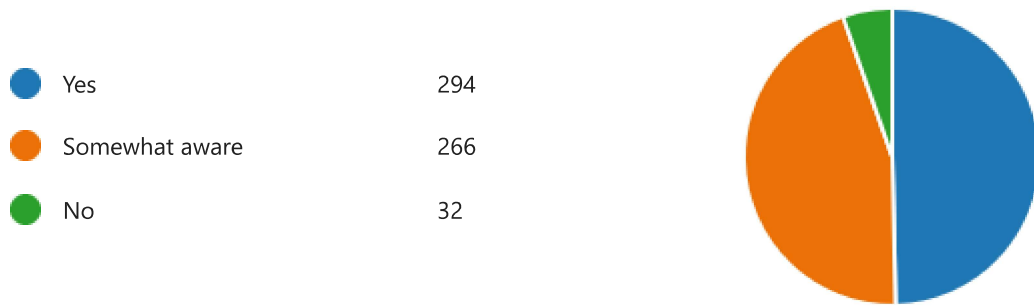
15. Below are some examples of online activities.

Which age verification method would you prefer to safeguard your child for each online activity?

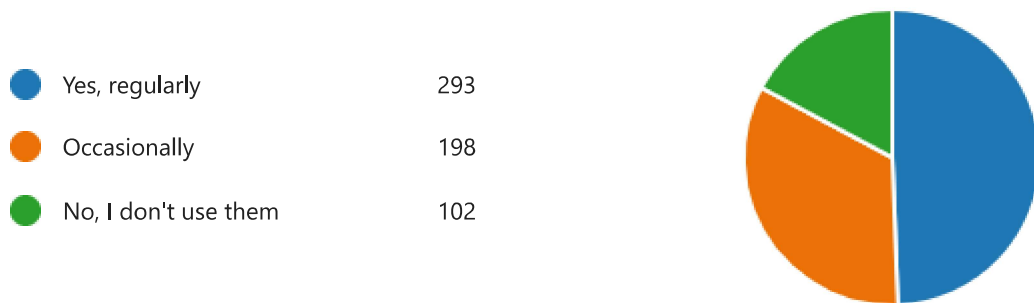
■ Age Assurance
 ■ Age Estimation
 ■ Age Gating
 ■ No age verification needed



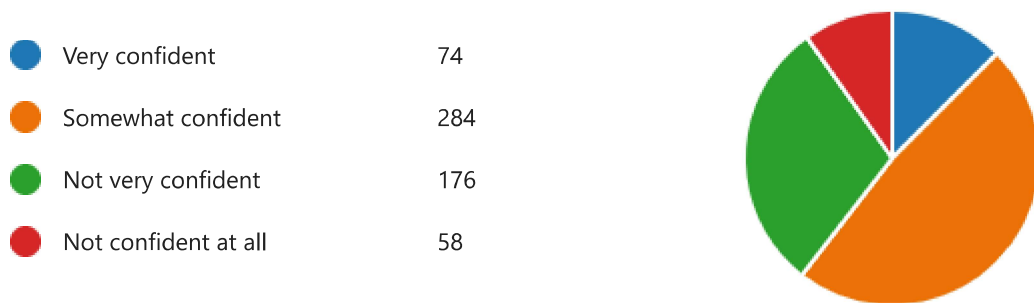
16. Are you aware of parental control features available on digital devices and online platforms?



17. Do you currently use parental control tools to monitor and regulate your child's digital usage?

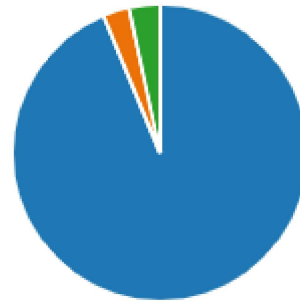


18. How confident are you in your ability to use parental control features to manage what content your children can access?



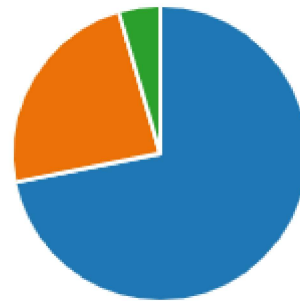
19. In your opinion, should parental controls be 'turned-on' by default for accounts of minors or where age is not verified?

- Yes, they should be turned on b... 554
- No, they should not be turned o... 17
- Not sure 20



20. Are you aware that the **content feed** and **advertisements** associated with video content are different from person to person based on their online activity?

- I am well aware that content fee... 427
- I have limited awareness but I'm... 140
- I was not aware 27



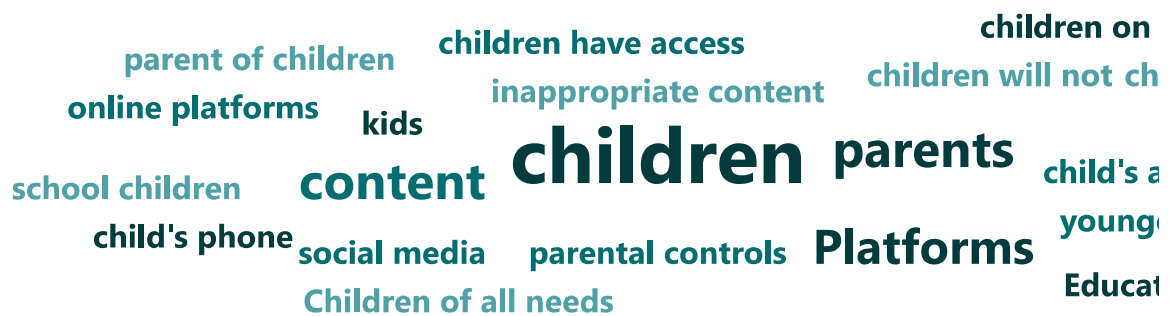
21. Keeping in mind the above survey, is there anything else the Commission for Online Safety could do to support children and young people with additional needs using video sharing platforms (Such as Tiktok, Instagram, YouTube)? (Parental Controls, Accessibility, Online Harms etc)

277

Responses

Latest Responses

113 respondents (41%) answered **children** for this question.



Online Safety Code Call for Inputs – Technology Ireland view

Technology Ireland’s response to Coimisiún na Meán’s Call for Inputs on developing Ireland’s first Online Safety Code.

Technology Ireland, the Ibec group representing the technology industry, welcomes the opportunity to respond to Coimisiún na Meán’s (CnaM) Call for Inputs on developing the first Online Safety Code for Video-Sharing Platform Services (VSPS). We are submitting the following points to highlight our views on the Code. As a sector we strongly voice our support for the protection of all users, including children and young people, from harmful online content, through codes and policy.

Ireland is a major hub for the technology sector in Europe, and the Irish Government has set out, through the National Digital Strategy (NDS), the ambition to “be a centre of regulatory excellence in Europe where both industry investments and European consumers are the winners”. Technology Ireland and its members strongly support this ambition. The NDS also notes that, “as the digital regulatory landscape becomes increasingly complex ... it will be more important than ever to ensure Ireland has a coherent regulatory framework”. CnaM is a key element of this framework and has an important role in creating and maintaining this coherence, in particular through careful alignment between a number of overlapping elements of EU and Irish law.

While this first Code is focused on the requirements for VSPS under the Audiovisual Media Services Directive (AVMSD), it is a precursor to further codes under the Online Safety and Media Regulation Act (OSMR), and so should be consistent and coherent with the requirements of the Digital Services Act (DSA). This first Online Safety Code should complement existing EU frameworks, ensure that the AVMSD is fully transposed in Ireland, and avoid the creation of new and potentially contradictory requirements.

We welcome CnaM’s intention to design the Code to minimise the potential for conflict and maximise the potential for synergies with the DSA. We strongly agree with this sentiment and the need to ensure that the Code does not conflict with the DSA (nor any other EU regulatory regimes).

To this end, when drafting the Code, CnaM should keep in mind the importance of the uniform application of the DSA’s harmonised rules to “put an end to fragmentation of the internal market” and “ensure legal certainty” (see Recital 4 DSA) and that Member States not adopt national measures dealing with requirements addressing the dissemination of illegal content online, as this is expressly recognised as an area which should be “fully” harmonised under the DSA (see Recital 9 DSA).

The Online Safety Code should be evidence based, proportionate and founded on research — either conducted or commissioned by CnaM or drawing upon the large body of research conducted by other countries. Online harms are global by nature and so any prioritisation should take this into account.

The Online Safety Code should operate with guiding principles and not be overly prescriptive, like other codes of practice. Having a principles based and outcomes focused approach allows for flexibility in solutions, and the ability for VSPS to iterate their products, safety features and solutions as new developments occur and based on the specific risks posed on each individual service. The Digital Trust and Safety Partnerships Best Practice Framework is one example of an approach that is clear, workable and fit for purpose, to which CnaM may wish to refer. Where possible, we recommend that CnaM seek to draw on existing codes and to work with the global regulatory community to support the harmonization of requirements.

CnaM may wish to consider a code based around the following guiding principles which can help shape a risk-proportionate and flexible approach such as recognising the transnational nature of the internet, to promote safety and systems based on best practice standards.

The AVMSD, and its implementing legislation, the OSMR, provide CnaM with a full suite of powers. Not every type of potential harm envisaged by the AVMSD requires the imposition of a binding online safety code. The Directive explicitly encourages the use of self- and co- regulation where appropriate, to meet the Directive's requirements. Supplementary non-binding guidance can also ensure greater adaptability and future-proofing, as new technologies and approaches (e.g. to content moderation) develop. These broader regulatory tools should be fully embraced by CnaM in the exercise of their regulatory role.

More generally, the operation of content limitation notices should be in line with that of the DSA and its equivalent provisions with consideration given to the territorial scope of such notices. Under the DSA, Member State authorities can, using content limitation notices, order the removal of illegal (and not merely harmful) content, whereas CnaM has power to require designated services to remove or restrict access to both illegal content and 'legal but harmful' content. Cooperation between CnaM and designated online services can be enhanced by providing scope for service providers to voluntarily suggest appropriate parameters to implement content limitation notice provisions, allowing for the potential innovation of self-regulating tools which would achieve the same objective.

The regulatory framework provided by AVMSD envisions an evolving and iterative process. Among the objectives of the Online Safety Code must be to establish a baseline of what existing measures VSPS have in place to meet the requirements of 28(b)1 and how effective these systems and processes are in mitigating risk. The AVMSD was drafted with an appreciation that not all VSPS will be identical. VSPS won't provide their services in the same way; nor will they meet the requirements of 28(b)1 in exactly the same way. The AVMSD also recognises that service providers may innovate and provide for new solutions that are not accounted for in the Directive. It is important that the first Online Safety Code respects this as well - for example, every VSPS should not be required to have every measure described in 28(b)2 in place, provided they can demonstrate that they are meeting the Directive's requirements using other measures or are otherwise achieving the required outcomes.

We also welcome clarification that this strategy will inform codes and guidance relating to the regulation of services in scope of the AVMSD and services designated under the OSMR Act only. We note that CnaM will separately acquire powers under the DSA later this year which will apply to a far broader group of providers, many of whom are neither providers of audiovisual media services nor designated services under the OSMR. We ask that CnaM undertake a fresh and separate consultation on its strategy with respect of DSA implementation to maintain focus on the specific requirements of the DSA and to avoid any inappropriate extension of measures and practices that are unique to the AVMSD and/or OSMR.

The reputational risk for Ireland cannot be underestimated with the successful rollout of the Online Safety Code. CnaM will be the lead regulator for many/most Technology Ireland members and maintaining good relations/information flows with regulators from other EU member states will be a key component of this leadership role, and CnaM should position itself to offer stable guidance in this regard.

Technology Ireland appreciates CnaM's extension to the period given for interested parties to respond to this consultation. To allow for thorough and constructive engagement, and in line with EU Better Regulation principles, we would ask CnaM to provide a minimum of six-weeks for future consultations. Additionally, we would ask that consultation requests avoid traditional holiday periods as far as possible (or with an additional extension of time), to ensure that the appropriate experts can provide input.

Technology Ireland looks forward to responding to further consultations on the regulation of online services operated by our members and welcomes future engagement with Coimisiún na Meán.

Online Safety: Developing Ireland's First Binding Online Safety Code for Video-Sharing Platform Services - Response to a Call for Input

Dr. Susan Leavy^{1,3}, Dr. Ruihai Dong^{2,3}

¹*School of Information and Communication Studies
University College Dublin, Dublin, Ireland*

²*School of Computer Science
University College Dublin, Dublin, Ireland*

³*Insight SFI Center for Data Analytics
University College Dublin, Dublin, Ireland*

Abstract

This submission focuses on ensuring that the effects of Video-Sharing Platform Services (VSPS), particularly on vulnerable users including children can be monitored and assessed effectively. Our submission identifies the key requirements to ensure transparency and oversight of algorithms. We identified the most relevant questions in the call for input and outline the access to algorithms and data that are required to conduct effective audits and ethics evaluations.

Introduction

The dissemination of content on Video-Sharing Platform Services (VSPS) is largely governed by recommendation algorithms. Given that the goal of these algorithms is primarily to maintain and increase user engagement, any associated social cost must be evaluated and continuously monitored. Our submission centres on ensuring that sufficient sufficient access to algorithms and data is mandated within the online safety code to enable effective oversight, transparency and auditing of VSPS.

Risks associated with VSPS stem from how users are categorised according to their online behaviour and content recommended that is deemed most likely to maintain their attention. Unfortunately, content that keeps people engaged can often cause harm, particularly to vulnerable groups[1]. Decisions on what content to recommend to whom are made by deep learning algorithms without human oversight making it difficult to identify risks.

The way recommendation algorithms suggest content has been shown to influence people's opinions, preferences and behaviour [11, 5, 5, 13]. This gives them potential to cause large-scale shifts in societal opinion. Political polarisation and effects on youth mental health for instance have been linked to how recommender systems disseminate content [3, 4]. In this submission we identify the questions that are most pertinent to the central focus on our response to this call for input.

Call for Input Responses

Question 1: What online harms should the code address?

Content distributed on VSPS has been shown to negatively influence people's preferences and behaviour [11, 12]. The most damaging effects are demonstrated in cases where tragic events such as suicide were linked to content recommended by AI Algorithms [6, 7, 8]. However, there is now increased awareness of how large-scale societal trends can be caused by recommender algorithms on VSPS. Such trends include, for instance, a rise in misogynistic views among boys, issues pertaining to youth mental health and political polarisation.

We propose therefore that the main priorities would be to ensure the protection of people's fundamental human rights with a focus on security, freedom of thought, privacy and freedom from discrimination. Humans should always be in control of how content is disseminated through-out a society. At present deep-learning algorithms have the ability to decide what content appears on a platform, presenting a profound risk to society.

Question 2: What types of online harms do you think should attract the most stringent risk mitigation measures by VSPS? How could we evaluate the impact of different types of harms e.g. severity, speed at which harm may be caused? Is there a way of classifying harmful content that you consider it would be useful for us to use?

The dissemination of illegal content should present the most stringent risk mitigation measures. Content that leads to the most damaging effects on society and individuals and categories of content that are inappropriate based on a person's age should also be a focus of stringent risk mitigation measures. Along with VSPS classifying and filtering harmful content, we propose that the code specifies access to recommender algorithms and data to ensure that auditors and regulators can verify the filtering of this content.

While there has been focus on identifying illegal and identifying clearly harmful pieces of content, there is an issue regarding VSPS recommending too much of the same kind of content to users. For instance, while one single video per week concerning weight-loss or depicting violence may not be damaging to a user, if this comprises 70 percent of what a users sees, then it may have a damaging effect. We propose that sufficient access to algorithms is mandated in order to test the categories of content that are being distributed to different user groups. Particularly for children, in order to protect them it is essential that regulators have the ability to present a societal level view of the kind of content that children are watching on VSPS. There are many existing readily available content classifiers that can classify content according to categories such as news, sports, health, technology, gaming, gambling, violence, weapons, profanity, mature content, alcohol and drugs. These kinds of classifiers can be an effective way of providing a report to government on what categories of content is being disseminated on their platforms.

Existing commercially available content filters can very effectively classify and filter website content on internet browsers. However they do not have access to content within VSPS such as YouTube and SnapChat. We propose that the Online Safety Code mandates that VSPS, especially

those used by children are mandated to allow parental control applications classify and filter content.

Question 3: Do you have reports, academic studies or other relevant independent research that would support your views? If you do, please share them with us with links to relevant reports, studies or research.

There are several reports that explore this topic and support our views. These are referenced throughout this document and are listed below in the references section. The most relevant is a recent report commissioned by the European Parliament on the effect of social media on teen mental health [1]. The other most pertinent report is the that by the Report by the Ada Lovelace Institute that outlines approaches to auditing VSPS [2].

Question 7: To what extent, if at all, should the Code require VSPS providers to take measures to address content connected to video content?

Comments in video have been known to be potentially harmful and should be classified as content and the same measures to ensure safety applied. For example, YouTube turned off comments on videos that feature children in 2019 due to the prevalence of predatory comments. While, for children's accounts on YouTube for those under 13 are not available, children between 13 and 18 remain unprotected. We propose that the online safety code for VSPS extend the protections that YouTube has extended to children under the age of 13, to children under 16 which is the digital age of consent in Ireland. Given the known risks of comments on videos therefore, obligations should be placed to provide highly accurate risk assessments and filtering of comments including text and image-based comments or otherwise remove the visibility of comments for children.

Question 11: What requirements should the Code have in relation to content rating? What do you consider to be current best practice? What experiences have you had using content rating systems on platforms and do you think they have been effective? What steps could we ask VSPS to take to ensure content is rated accurately by users?

We propose a new approach to rating of content that departs from traditional age rating approaches. If it is mandated that all videos are automatically classified according to certain categories, then categories of content deemed inappropriate could then be filtered out based on age. For instance, videos depicting extreme violence could be classified and made unavailable for children. This would allow content relating to gambling or violence for example, to be classified and automatically filtered.

Method	Description
Code Audit	Auditors analyse VSPS code directly
User Survey	Surveys/interviews are conducted to understand users' experience
Scraping Audit	Auditors scrape data from a platform automatically
API Audit	Auditors access algorithms through an API
Sock Puppet Audit	Auditors set up user profiles and simulate user behaviour
Crowd-Sourced Audit	Real users provide information on their experience through manual or automated reporting

Table 1

Adapted from Technical methods for regulatory inspection of algorithmic systems, Ada Lovelace Institute(2021).

Question 12: What requirements should the Code have in relation to parental control features? How can we ensure that VSPS providers introduce the mechanism in a user-friendly and transparent way? Can you point to any existing example of best practice in this area? Should parental controls be 'turned-on' by default for accounts of minors or where age is not verified?

Video sharing applications to be manageable by one single parental control application. These applications should also be designed and managed by a third-party provider to ensure trust, transparency and accessibility. Many existing parental security applications are effective for some applications but are locked out of accessing content in many popular VSPS and are unable to provide parental oversight or security. For users who are minors, providing parental controls by default should be mandatory and settings should filter age-inappropriate categories of content by default.

Question 18: What approach do you think the Code should take to risk assessments and safety by design? Are there any examples you can point us towards which you consider to be best practice?

Risk assessments may be conducted through audits of algorithms and data on VSPS. This can be done using a variety of established methods. However, once increased access to data and algorithms are mandated it will be possible to develop more effective and accurate automated audits and monitoring tools. The following classification of audit approaches was outlined by the Ada Lovelace Institute 1. Access to platforms in order that third-party auditors can conduct these audits must be mandated.

Question 20: What approach do you think we should take in the Code to address feeds which cause harm because of the aggregate impact of the content they provide access to? Are there current practices which you consider to be best practice in this regard?

We propose a new approach that focuses on classifying content and monitoring societal trends in content distribution. This can be done by creation of a third-party system that monitors

content dissemination patterns. Alternatively, logs can be generated by companies and analysed by a third partner to create a picture of aggregate content dissemination patterns.

The key to enabling the analysis of the impact of how content is disseminated to user groups is the ability to identify the category of content. There have been significant recent advances in online content classification using deep learning-based techniques. An effective solution involves using deep neural networks to encode the online content into dense embeddings, and feed these into fully connected layers for classification. For example, pre-trained language models can be used to learn embeddings for textual data and CNN-based networks can be used to generate representations for images and videos [10, 9]. This project will build on this work and incorporate these approaches in identifying potentially harmful online text or video content. Such analysis can be conducted by third party auditors, but this relies on them having access to algorithms and data or access through an API.

Question 24: What is the significance of safety code, what potential advantages does it bring to Ireland, and why is it timely to focus on this matter now?

The significance of a safety code for VSPS lies in enhancing user security, content moderation, and data protection. Implementing such a code can provide Ireland with several potential advantages, including bolstering its reputation as a safe and responsible digital hub, attracting tech investments and businesses, and fostering a safer online environment for its citizens. It is timely to focus on this matter now because of the growing importance of online platforms and the increasing need to address issues like harmful content, privacy breaches, and cyber threats.

There is also a lot of state of the art research being conducted in Ireland on the development of Trustworthy AI. Ensuring access to algorithms and data either directly or through APIs, will ensure that Ireland is a world leader in developing methods and techniques to conducting rigorous ethical audits of AI systems.

1. Conclusion

References

- [1] Kirsty Park, Debbie Ging, Cian McGrath, and Shane Murphy. 2023. The impact of the use of social media on women and girls, Study Commissioned by EU Parliament.
- [2] Ada Lovelace Institute 2021. Technical methods for regulatory inspection of algorithmic systems. <https://www.adalovelaceinstitute.org/report/technical-methods-regulatory-inspection/>.
- [3] Hunt Allcott, Luca Braghieri, Sarah Eichmeyer, and Matthew Gentzkow. 2020. The welfare effects of social media. *American Economic Review* 110, 3 (2020), 629–76.
- [4] Monica Anderson, Jingjing Jiang, et al. 2018. Teens, social media & technology 2018. *Pew Research Center* 31, 2018 (2018), 1673–1689.
- [5] Henry Ashton, Matija Franklin, Rebecca Gorman, and Stuart Armstrong. 2022. Recognising

the importance of preference change: A call for a coordinated multidisciplinary research effort in the age of AI. AAI.

- [6] Coroner's Service UK. 2022. REGULATION 28 REPORT TO PREVENT FUTURE DEATHS. (2022).
- [7] Gonzalez v. Google LLC. 2022. <https://www.oyez.org/cases/2022/21-1333>.
- [8] US Sentate. 2021. Senate Hearing Transcript: Facebook Whistleblower Frances Haugen Testifies on Children & Social Media Use..
- [9] Sun, C, Qiu, X, Xu, Y, and Huang, X. 2019. How to Fine-Tune BERT for Text Classification. In *Proceedings of the 2019 CCL Conference*. 194-206.
- [10] Wang, J. et al.. 2019. CNN-RNN: A Unified Framework for Multi-label Image Classification Preprint at <https://doi.org/10.48550/arXiv.1604.04573>.
- [11] Micah D Carroll, Anca Dragan, Stuart Russell, and Dylan Hadfield-Menell. 2022. Estimating and Penalizing Induced Preference Shifts in Recommender Systems. In *International Conference on Machine Learning*. PMLR, 2686–2708.
- [12] Charles Evans and Atoosa Kasirzadeh. 2021. User Tampering in Reinforcement Learning Recommender Systems. *arXiv preprint arXiv:2109.04083* (2021).
- [13] Holly B Shakya and Nicholas A Christakis. 2017. Association of Facebook use with compromised well-being: A longitudinal study. *American journal of epidemiology* 185, 3 (2017), 203–211.



Laura Forsythe
Coimisiún na Meán,
2 – 5 Warrington Place,
Dublin D02 XP29

VSPSregulation@cnam.ie

September 4, 2023

Dear Ms Forsythe,

Thank you for the opportunity to respond to your consultation as you prepare to implement the Audio-Visual Media Services Directive and wider online safety legislation for video sharing and other online platforms established in Ireland.

VerifyMyAge and *VerifyMyContent* were created by a team of eCommerce specialists that understand the importance of robust and effective age verification and content moderation solutions, with the aim to reduce the access children have to age-restricted goods, content and services.

Our submission outlines the services we currently offer to provide you with evidence of what is already possible, and widely used in live operation by our existing clients around the world.

We work with some of the largest digital platforms, while also providing easy-to-apply plug-ins for all the major e-commerce marketplaces, enabling compliance whatever the size of the business.

Our age assurance solution aligns with the age and developmental categories introduced by the UK ICO. We are also certified by the Age Check Certification Scheme as being PAS1296 compliant, as per the Code of Practice for Online Age Verification Service Providers in the UK. We are also on the UK government's Directory of UK Safety Tech Providers.

Question 9: How should we ask VSPS providers to introduce and design a flagging mechanism in the Code? How can we ensure that VSPS providers introduce the mechanism in a user-friendly and transparent way? How should we ask VSP Providers to report the decisions they've made on content after it has been flagged? To what extent should we align the Code with similar provisions on flagging in the DSA?

Unless a site has low enough volumes of content uploaded to allow for manual review, we believe it is essential to deploy AI-based moderation and reporting tools.

We leverage a global network of content moderation experts, artificial intelligence and machine learning to review and moderate user-generated content prior to publication.

Real-time and continuous live stream moderation is possible and highly effective, using artificial intelligence, machine learning networks and global content moderation. Stop requests can be made within seconds of a live streaming offence being committed.

A team of human moderators review all content flagged for review during AI content moderation, and if required, by users. In addition to this, as part of our ongoing quality control process, this team moderates a sample set of all content not flagged by AI or users.

Question 10: What requirements should the Code include about age verification and age assurance?

As specialists in age assurance and compliance, we utilise methodologies independently certified as meeting the requirements of PAS 1296:2018 – Code of Practice for Online Age Verification.

Once a customer is verified with VerifyMyAge, they stay verified across any VerifyMyAge-integrated platform allowing for re-usability of existing age checks.

VerifyMyAge operates as the layer above deep-tech, creating an age assurance ecosystem of data, deep tech and artificial intelligence partners, resulting in the most comprehensive and effective age verification and estimation solution available.

VerifyMyAge provides age assurance (age verification and age estimation) while ensuring customers are not distracted from their purchase or online experience. Our solution can be integrated either directly - within the browser post-checkout - or indirectly through email or SMS.

We have two main methods of providing age assurance and identifying the age or age range of an online user.

1. Age verification allows us to provide higher levels of certainty of the age or age-range of a user e.g. ID scan, credit bureau check, mobile phone number check, credit card check
2. Age estimation, as the name suggests, provides an estimate of a user's age or age range and often relies on artificial intelligence and machine learning techniques e.g. facial age estimation, voice age estimation, email address check

Question 10: What sort of content should be shown by default to users who are logged out or in private browsing mode and whose age cannot be verified or assured?

The Irish “Fundamentals” Code applies largely to any business or “information society services likely to be accessed by children”, that has either been established in Ireland or targets Irish users from outside the EU. This will be the most influential piece of regulation when considering users who are not logged into an account or cannot be recognised as previously verified thanks to private browsing.

Companies that operate streaming services, social media sites, or provide online gaming, video, films, music, education technology or connected toys will be the most affected. The “Fundamentals” will also impact online retailers and require the need for age appropriate retargeting onsite.

These standards provide built-in protection to allow children to explore, learn and play online, ensuring their best interests are the primary consideration of designers and developers for online services so it is important the Commission reminds platforms these apply across the board.



Question 10: What evidence is there about the effectiveness of age estimation techniques?

We strive to provide an exhaustive choice of available age verification and estimation methods and data-sets. A combination of research and our own data shows this leads to a higher rate of successful age verifications and customer satisfaction. This is why over 99.9% of our verifications of adults are successfully approved.

Database Check

We capture information from customers' orders, and verify their age without them having to act.

Mobile Phone Number

By sending a text to a UK mobile number, we verify the phone is authorised for use by a person aged 18+.

Facial Biometrics

Using AI-powered age estimation, we'll have a customer take a short selfie video, to indicate whether they are aged 25+ (or any other required threshold)

Government Issued ID

We verify customers' age and identity using their scanned government ID or driver's license.

Credit Card

We verify the cardholder is 18+ using their credit card records.

Open Banking

We'll verify a customer's age by accessing their main bank account through a secure API connection.

Email Address

We'll use a range of data points to determine a customer's minimum age using their email address. This form of estimation does not require any biometric data.

VerifyMyAge is designed to be as efficient and seamless as possible. To minimise friction and ensure positive customer experiences, where possible, we attempt to authenticate/estimate a customer's age using the PII already provided to our clients, via data partners, subject of course to GDPR requirements for a clear legal basis on which to process this data.

Question 10: What current practices do you regard as best practice?

We have developed an Age Assurance Ecosystem, allowing us to offer best-in-class solutions across a range of technologies.

VerifyMyAge works with a comprehensive range of technology partners to create an age assurance ecosystem of data, deep tech and artificial intelligence partners, resulting in the most comprehensive and effective age verification and estimation solution available. We regard our work with these partners as best practice:

Experian



With Experian's comprehensive data resources, VerifyMyAge can seamlessly verify information submitted during the verification process.

Facetec

The world leader in 3D Face Liveness & Matching software. Enabling quick and simple age estimation and document face matching.

Yes

Yes enables the sharing of open banking data to assess age and identity information.

Schufa

With Schufa's comprehensive data resources, VerifyMyAge can seamlessly verify information submitted during the verification process.

Amazon Rekognition

Amazon's computer vision platform enables VerifyMyAge to increase the accuracy of age estimation technologies.

Question 11: What requirements should the Code have in relation to content rating? What do you consider to be current best practice? What experiences have you had using content rating systems on platforms and do you think they have been effective? What steps could we ask VSPS to take to ensure content is rated accurately by users?

The use of VerifyMyAge's PAS 1296:2018 Certified approach is considered by us, if properly implemented by the client, to be objective evidence of conformity with the AADC requirements for Age Assurance. This would add significant value to services applying the 15 principles of the AADC."

Tony Allen, Chief Executive Officer at Age Check Certification Scheme.

Question 12: What requirements should the Code have in relation to parental control features? How can we ensure that VSPS providers introduce the mechanism in a user-friendly and transparent way? Can you point to any existing example of best practice in this area? Should parental controls be 'turned-on' by default for accounts of minors or where age is not verified?

Our age estimation methods can be used to help children safely access content online with a parent or guardian's consent.

When a child attempts to access services, they will be asked to provide contact details for a parent or guardian. The parent or guardian will be asked to verify their age and give consent for their child to access the content.

Question 15: How should we ask VSPS providers to address content moderation in the Code? Are there any current practices which you consider to be best practice? How should we address automated content detection and moderation in the Code?

We understand the challenges a user-generated content platform faces, so we've created the complete end to end identity verification and content moderation solution, providing our clients and their users with the tools to ensure a safe and trustworthy online experience for everyone. This is a complete user authentication, content moderation, and complaint



management solution for user-generated content platforms - making the internet safer for all.

VerifyMyContent allows online services to streamline the entire verification process, from uploader verification to consent and content moderation, without compromising on security. This allows services to get their content online fast.

Using a combination of AI and human moderators ensures content is moderated and published as quickly as possible. We can block offensive content because it is automatically flagged for urgent review via the content moderation dashboard, putting services in control of their complaints procedure.

1. Age and identity verification.

Documented age and identity verification for all people depicted and those uploading the content.

2. Content moderation.

Content review process prior to publication.

3. Complaint resolution.

Complaint resolution process that addresses illegal or non consensual content within seven business days.

4. Appeals process.

Appeals process allowing for any person depicted to request their content be removed.

User-generated content platforms face a growing number of trust and safety risks, and the need for robust compliance and safeguarding has never been greater.

For many clients, content moderation is already a high priority as it is required to protect their revenue, as a result of Mastercard's AN5196 regulations (adult entertainment sector). Compliance with these is essential to prevent any potential disruption to the processing of payments arising if services breach these rules, implemented by the acquiring banks on behalf of Mastercard.

We supply a powerful and innovative combination of identity verification, automated and manual content moderation, quality control and reporting, giving our clients the complete end to end compliance and safeguarding solution for their business.

This video illustrates how we achieve this: <https://vimeo.com/610652354>

To secure compliance with these requirements, we offer both content provider & participant Verification. This can automate the entire process of verifying the age and identity of content providers and participants using a combination of AI and identity document scanning technologies. Our content moderation solution works over a vast range of media types, ensuring maximum compliance and user safety whatever the platform, including:

- Video
- Live Streaming
- Images
-

Question 16: What requirements should the Code include about procedures for complaint-handling and resolution, including out-of court redress or alternative-dispute resolution processes? To what extent should these requirements align with similar requirements in the DSA? What current practices could be regarded as best practice? How frequently should VSPS providers be obliged to report to the Commission on their complaint handling systems and what should those reports contain? Should there be a maximum time-period for VSPS providers to handle user complaints and if so, what should that period be?

With our VerifyMyContent product, user-reported content is flagged and reviewed by a dedicated content moderation team ensuring complaint resolution within seven working days. This could therefore be a reasonable benchmark adopted by the Commission in its regulations.

Question 17: What approach do you think the Code should take to ensuring that the safety measures we ask VSPS providers to take are accessible to people with disabilities?

As described above, we offer a wide range of options for users when verifying their age in order to promote accessibility, achieving a very high pass rate across a diverse population of users.

Question 19: How do you think that cooperation with other regulators and bodies can help us to implement the Code for VSPS?

We have experience of seeking approval under Article 42 of GDPR and comment on this form of co-regulation. Where possible, national regulators should accept mutual recognition of one another's certification programs.

Question 22: What compliance monitoring and reporting arrangements should we include in the Code?

For our VerifyMyContent product we offer clients Real-Time Monitoring. With the power of artificial intelligence, we continuously monitor live streams for potential violations, reducing the risk of harm to performers and users.

Real-time data is consolidated into easily accessible, automatically generated monthly reports detailing all compliance actions taken.

These reports can be configured to meet any specific regulatory requirements provided the underlying data has been captured, and subject to data minimisation and privacy-by-design principles.

Automated monthly reporting gives our clients quick and easy access to compliance reports, enabling them to instantly demonstrate to regulators such as the Commission all compliance measures undertaken.

Question 23: Should the Code have a transition period or transition periods for specific issues? Which areas touched on in this Call for Inputs may VSPS providers require time to transition the most? What time frame would be reasonable for a transition period?

As described above our solutions are offered either as simple plug-ins to established platforms, or through APIs, or through bespoke integrations.

- Plug-in solutions can usually be adopted with a few hours of effort to complete the set-up process.



- API integration may require a few days including the necessary integration testing.
- Bespoke integration is completely dependent on the range of methods, the volume of checks, and the extent of testing clients require, but can be achieved within one or two technical sprints over, typically 4-6 weeks.

We would be pleased to provide demonstrations of any of our capabilities to assist the Commission in understanding what is currently available in the market, to businesses of all sizes, with minimal effort and time required to implement our solutions.

Yours sincerely,

Andy Lulham

Chief Operations Officer
VerifyMy

Coimisiún na Meán: Online Safety Code Consultation Response

September 2023

About WeProtect Global Alliance

WeProtect Global Alliance ('the Alliance') is a non-profit that brings together people and organisations with the knowledge, experience and influence to transform the global response to child sexual exploitation and abuse online. As of August 2023, its membership is comprised of 102 government [members](#) – including the Irish Government – 66 companies, 92 civil society groups and 9 international organisations.

The Alliance supports Coimisiún na Meán ('the Commission') in initiating the first steps to develop Ireland's first binding Online Safety Code for Video-Sharing Platform Services (VSPS). This code will help the Commission to deliver one of its three key goals, namely setting up a new regulatory regime for online safety. Below the Alliance is submitting feedback on the guiding questions shared by the Commission in its [Call for Inputs](#) entitled [Online Safety – Developing Ireland's First Binding Online Safety Code for Video-Sharing Platform Services](#), guided by the [consultation document](#) and the [consultation guidelines](#).

As a multi-sector membership organisation spanning governments, civil society, the private sector and international non-governmental organisations, WeProtect Global Alliance occupies a unique position in the child protection sector and thus has a comprehensive viewpoint of both the threat landscape for children in a digital environment and the current response to child sexual abuse online. As a consequence, the Alliance understands the role and importance of codes that improve online platform regulation as part of the wider response to tackling child sexual abuse and exploitation online; one which involves coordinated, consistent and strategic action by a range of stakeholders – as set out in our well-established [Model National Response](#) (MNR) framework.

Responses to relevant questions set out in *Call for Inputs: Online Safety – Developing Ireland's First Binding Online Safety Code for Video-Sharing Platform Services*

Question 1: What do you think our main priorities and objectives should be in the first binding Online Safety Code for VSPS? What are the main online harms you would like to see it address and why?

The most dangerous and severe harms, that have the potential to cause real and significant emotional, physical and social harm, both immediately and in the longer term, should be prioritised. The harmful online content relating to 42 criminal offences under Irish law listed in

Schedule 3 of the 2009 Act as amended would fall under this category. WeProtect Global Alliance believes that child sexual abuse should be a top priority for the Commission, as well as the non-consensual sharing of intimate images, child and human trafficking and domestic violence.

Children today face a sustained threat of child sexual exploitation and abuse online. WeProtect Global Alliance's latest [Global Threat Assessment](#) shows that the scale and complexity of this threat are increasing. Europe is also home to a big portion of the abuse, with [over 66% of the URLs containing child sexual abuse material hosted on servers based in the EU](#) in 2022. Children and young people are experiencing online sexual harm across the European Union and the speed and scale in which they experience harm after getting online is particularly concerning. [Our recent study, conducted by Economist Impact](#), explored the experiences of 2,000 18-year-olds across four European countries to better understand their exposure to online sexual harms during childhood. It found that almost 7-in-10 of those surveyed (68%) experienced at least one form of online sexual harm during childhood. Given that the average class size across the European Union is approximately 20 students, 13 to 14 (13.6) children in that class will statistically be affected by online sexual harms before their 18th birthday.

The nature of online harm has continued to grow and diversify at an unprecedented rate – and video-sharing platform services lie at the heart of much of this harm. [Research by Ofcom](#) shows that 70% of VSPS users encountered exposure to at least one potentially harmful online experience on such platforms within the three months prior to research being conducted. Evidence indicates an increase in:

- the incidence of online grooming;
- the volume of child sexual abuse material available online;
- the sharing and distribution of child sexual abuse material; and
- Livestreaming of abuse for payment.

The Alliance's [Global Threat Assessment 2021](#) states that “the best opportunity for change is to improve online safety for children and reduce opportunities for offenders”.

Question 2: What types of online harms do you think should attract the most stringent risk mitigation measures by VSPS? How could we evaluate the impact of different types of harms e.g. severity, speed at which harm may be caused? Is there a way of classifying harmful content that you consider it would be useful for us to use?

Illegal and harmful content, where the severity and immediate impact of the harm is most dangerous, or that poses an imminent threat to the lives and safety of vulnerable users, should attract the most stringent risk mitigation measures by VSPS. Child sexual abuse content is a serious crime that can have devastating emotional, social and physical consequences for victims and survivors.

Video sharing platform services should have proactive, robust and comprehensive risk mitigation measures in place to detect, report, block and remove this content, as well as to report it to the authorities in order for offenders to be brought to justice. In explaining the risk mitigation measures adopted, tech companies should be able to show that user safety was prioritised in product design and engineering decisions.

We believe it is important to have a common, internationally recognised and accepted, typology for identifying and responding to online harms, e.g., EU Kids Online (2009) 3 C's framework, which classifies online risks to children into content, contact and conduct categories. It can facilitate multistakeholder discussions—which are important for a multifaceted issue such as online harms—and contribute to the development of policy interventions consistent across different regulatory regimes.

Question 3: Do you have reports, academic studies or other relevant independent research that would support your views? If you do, please share them with us with links to relevant reports, studies or research.

[2021 Global Threat Assessment](#) – this report is our most comprehensive yet and shows how the global response to child sexual exploitation and abuse online needs a new approach. It explores many elements of the threat and particular threats relating to video-sharing platform services (e.g., disseminating child sexual abuse material, live-streaming child sexual abuse, non-consensual sharing of videos and images, and more). A 2023 Global Threat Assessment is set to be released in October 2023.

[Child 'self-generated' sexual material online: children and young people's perspectives](#) – in this research conducted with Praesidio Safeguarding, we listened to children and young people's views on the issue of 'self-generated' sexual material in three different country contexts – Ireland, Ghana, and Thailand.

[Estimates of childhood exposure to online sexual harms and their risk factors in the European Union](#) – this study, conducted by Economist Impact, explores the experiences of 2,000 18-year-olds across four European countries (France, Germany, the Netherlands and Poland) who had regular access to the internet as children to understand their experiences of and exposure to online sexual harms during childhood.

[The role of age verification technology in tackling child sexual exploitation and abuse online](#) – this intelligence briefing explores the role that age assurance can play in safeguarding children, the current regulatory landscape around age and different methods of age assurance.

More resources regarding specific issues in the response to tackling child sexual exploitation and abuse online are available via our library: <https://www.weprotect.org/library/>

Question 4: What approach do you think we should take to the level of detail in the Code? What role could non-binding guidance play in supplementing the Code?

The language of the code needs to be clear and detailed when it comes to tackling online harms that we are already aware of and have the tools to reduce them. Non-binding guidance is a helpful in ensuring that the code is flexible and adaptable to emerging technologies and harms.

Question 5: What do you think would be the most effective structure for the Code? What are the most important factors we should consider when we decide how to structure the Code?

The code is an opportunity to create a sole holistic framework with consistent safety standards that ensure that all video sharing platforms operating in Ireland and used by Irish citizens are safe for all users.

Firstly, it will be necessary to establish a clear set of roles, responsibilities and expectations regarding the code, covering a range of different actors including the government, the regulator, industry, civil society, and the public to name a few. For example, the code should clarify if the regulator's primary functions will be protection, prevention or both. It should also specify if and how it plans to embed Safety by Design in evolving technological trends.

Secondly, the scope of the code will need to be clarified. Identifying the types of harm that will fall under the rules is an essential task and to be effective, the Code will need to set out priorities. There will need to be explicit and strict safety obligations on child sexual abuse and exploitation online, and for larger or high-risk video sharing platforms.

Thirdly, industry will have to assess their current response to the online harms listed in the code and then implement the necessary safeguards to ensure they are up to standard and tackling harm on their platforms (risk management, processes to prevent, remove, reduce, and block unsafe content, consumer-focused processes, transparency reports). The regulator should have appropriate powers to approve and enforce the Code, as well as challenge VSPS who do not comply with the standards established by the code.

Question 6: How should we design the Code to minimise the potential for conflict and maximise the potential for synergies in how platforms comply with it and the DSA?

The EU's Digital Services Act (DSA) has been a vital development in establishing new rules to regulate against harmful forms of online content such as disinformation, hate speech and illegal content, including child sexual abuse material. For the code to minimise conflict and maximise synergies, it is crucial to clearly delineate the DSA's requirements, which highlight several specific obligations for platforms, such as the removal of illegal content, the protection of minors, and the promotion of user safety.

The code should be designed to complement and reinforce the requirements of the DSA, while also not shying away from additional measures should the Commission feel that certain elements of online safety have been overlooked in the DSA (for example, stricter measures on harmful content). The Alliance is supportive of the DSA's risk-based approach and believe that this is a good foundation for the Code. Given that the online world is constantly evolving, and new challenges, harms and threats to safety are also unfolding, the Code should be designed in a tech neutral way to be flexible and adaptable, so that it can respond to new and emerging threats and harms.

Transparency and accountability are key tenets of the DSA and should also be central to the Code. Consulting, listening to and involving a wide range of stakeholders (e.g., platforms, users, regulators, and civil society organisations) will be crucial to ensuring that the Code is balanced and effective. Guidance for these stakeholders needs to be clear and enforcement measures need to "have teeth" to ensure that platforms comply with requirements. Approaches to regulation can vary, ranging from the 'lighter touch' approaches like voluntary codes of practice, good-practice guidance, and reputational incentives to more stringent or prescriptive measures

such as mandatory codes of practice, investigations, and even legal penalties or even criminal sanctions. Given that the voluntary codes and self-regulation of the recent years has not delivered the best results, more prescriptive measures should be explored, especially for VSPS who repeatedly fail to comply with the code. By following these recommendations, we can design a Code that is effective in protecting online safety and that is also fair and balanced.

Question 7: To what extent, if at all, should the Code require VSPS providers to take measures to address content connected to video content?

Content, such as captions, hyperlinks embedded in video, etc., connected to video content is a particularly challenging issue. WeProtect Global Alliance's recent [briefing on link-sharing and child sexual abuse](#) highlighted that one of the main challenges for many service providers is how to moderate links presented on their platforms through which users are taken to harmful and illegal content that is hosted on a different site. WeProtect Global Alliance's [2021 Global Threat Assessment](#) highlighted that there are signs of offenders moving away from the curation of personal collections of child sexual abuse material and preferring 'on-demand' access to content via the sharing of links that lead to child sexual abuse content. Links to files containing child sexual abuse content are posted across multiple sites and often used as part of offender-to-offender sharing. This creates a raft of challenges for law enforcement. Material is often published and hosted in different jurisdictions, which complicates evidence-gathering. There is little available data on how companies are responding, which makes it difficult to assess the efficacy of responses. The action taken by industry can depend on where the links take users. For example, a link may take a user to content hosted externally, or link to an image-hosting site or website, or to group chats on group messaging apps and forums. All these may be harmful yet require different responses.

Collaboration with leading safety technology organisations forms an essential part of the response for leading industry players. Many participants at the roundtable cited the Internet Watch Foundation's (IWF) [URL List](#) as a helpful tool in identifying potential harms and blocking access to illicit webpages and material. [Project Arachnid](#) in Canada is also an effective technology to combat link-sharing. It identifies child sexual abuse material by crawling specific publicly accessible URLs reported to [CyberTipline](#), as well as URLs on the surface web and dark web that are proven or known to host child sexual abuse material. It detects URLs that host media and matches content against a database of digital fingerprints. As soon as Project Arachnid detects a match in fingerprints, a removal notice is automatically issued requesting the hosting provider to take it down. It follows up on this request by recrawling URLs linking illegal content every day until the content is taken down.

The extent to which VSPS providers should be required to take measures to address content connected to video content is complicated. Measures identified in the Alliance's work on link sharing include creating as hostile an environment as possible for offenders and potential offenders, constantly innovating technology and increasing the deployment of artificial intelligence to respond to the scale and complexity of this particular harmful activity and increased collaboration between internet service providers, telecommunication companies, technology companies, safety tech, law enforcement authorities, security agencies, reporting centres, hotlines and victim support services.

Question 8: How should we ask VSPS providers to introduce a feature that allows users to declare when videos contain advertising or other type of commercial communications? Should the Code include specific requirements about the form in which the declaration should take? What current examples are there that you regard as best practice?

Declaring advertising needs to be clear to understand and easy to use for providers to declare when content contains advertising. Advertising on VSPS providers comes in many forms so it is important to encourage different types of disclosure for the different types of advertising. Policies on advertising need to be enforceable and action should be taken against users who violate the policy, such as removing their videos or suspending their accounts.

Question 9: How should we ask VSPS providers to introduce and design a flagging mechanism in the Code? How can we ensure that VSPS providers introduce the mechanism in a user-friendly and transparent way? How should we ask VSP Providers to report the decisions they've made on content after it has been flagged? To what extent should we align the Code with similar provisions on flagging in the DSA?

The DSA imposes new mechanisms allowing users to flag illegal content online, and for platforms to cooperate with specialised 'trusted flaggers' (accredited groups with expertise and experience in particular harms) to identify and remove illegal content. It is understood that priority channels are being created for trusted flaggers to report illegal content. Under the DSA, all platforms, except those with fewer than 50 employees and whose annual turnover and/or annual balance sheet total does not exceed €10M, are required to set up complaint and redress mechanisms and out-of-court dispute settlement mechanisms, cooperate with trusted flaggers, take measures against abusive notices, deal with complaints, vet the credentials of third party suppliers, and provide user-facing transparency of online advertising. Such flagging tools for users need to be accessible and easy to use for all users, especially specific groups, such as making tools child-friendly and disability friendly. Services providers need to provide users with different reasons to flag content such as child sexual abuse content, hate speech, violence, or privacy violations and they should also be able to provide additional comments, feedback or evidence to ensure that claims of harmful content can be as robust as possible. Users should also be able to contest decisions by VSPS if content flagged by them as inappropriate is not actioned. We should also be cautious with flagging that not too much onus is placed on users to report. Platforms have a big responsibility in hiring enough content moderators, training and supporting them well and ensuring that they have the tools and resources to act swiftly to identify and takedown harmful content themselves.

Question 10: What requirements should the Code include about age verification and age assurance? What sort of content should be shown by default to users who are logged out or in private browsing mode and whose age cannot be verified or assured? What evidence is there about the effectiveness of age estimation techniques? What current practices do you regard as best practice? Where accounts are not age verified should default privacy settings be used, should content default to universal content and should contact by others be more limited?

Just as we protect children offline – they can't freely walk into a nightclub or buy a bottle of wine – the same protections need to be implemented online. Whilst there are many positive opportunities available online, increasing numbers of children are accessing explicit content,

chatting to strangers or being coerced into sharing images of themselves. In the UK, [research shows](#) that many children - some as young as 7 years old - stumble upon adult pornography online, with 61% of 11-13-year-olds describing their viewing as mostly unintentional. WeProtect Global Alliance believes that age assurance is one of the tools that can be used to create digital products safe by design. Our 2021 [Global Threat Assessment](#) highlighted that age estimation and verification tools are some of the Safety by Design solutions with the most potential to reduce the risk of online grooming. Such technology is still relatively nascent but could be used to exclude predators from children's forums and ensure age-appropriate online experiences.

There are many different methods for carrying out age assurance checks, from more 'traditional' types such as ID, mobile phone number or credit card checks, to evolving technologies such as facial age estimation, identity apps and social media proofing. In order to ensure that users remain in control of their privacy, the Alliance believes that it is important to provide consumers with a choice as to which age estimation tools they use to confirm their age online.

More information is available in our [briefing on age estimation techniques](#).

Question 11: What requirements should the Code have in relation to content rating? What do you consider to be current best practice? What experiences have you had using content rating systems on platforms, and do you think they have been effective? What steps could we ask VSPS to take to ensure content is rated accurately by users?

Content rating online is complex. One potential way to tackle this would be to assess the intended audience of the content (children, adults, etc.) and the themes/subjects covered in the content itself. For example, when considering targeting content in mixed-age environments, content producers should be expected to use audience targeting tools to target content away from children. Important elements to consider are whether the content is factual, fictional, or a mix of both or if it is violent, sexual, or otherwise harmful. For ratings involving children, it is also important to bear in mind if the content is developmentally appropriate for their age. The context in which the content is being shared – between friends or strangers, in public or in private, etc. - is also important. Once these factors have been considered, it is possible to develop a rating system that is appropriate for VSPS.

Question 12: What requirements should the Code have in relation to parental control features? How can we ensure that VSPS providers introduce the mechanism in a user-friendly and transparent way? Can you point to any existing example of best practice in this area? Should parental controls be 'turned-on' by default for accounts of minors or where age is not verified?

Parental controls and content filters are also key tenets of the Safety by Design approach. As highlighted in our 2021 [Global Threat Assessment](#), many mainstream platforms already incorporate some of these, for example, gaming platform Roblox has built-in security software blocking explicit content and preventing young users sharing their contact information. Social networking platform TikTok has introduced default privacy and safety settings for under 18s. Instagram is adding safety features to protect teenagers from unwanted direct messages from adults they don't know. YouTube has developed 'Supervised Experiences' for children under 13, limiting their ability to upload content, chat or receive comments, and helping parents manage content they access.

In terms of requirements for parental control features, at the top level they need to be flexible, effective, easy for both parents and children to use, while also protecting and upholding the highest levels of user privacy possible. There are many different types of parental controls available on online services, but some of the most popular and effective include screen time limits, app blocking, web filtering, location tracking and activity reporting.

One of the problems with parental control features is that a lot of platforms have introduced them, but parents do not necessarily know they exist. Platforms need to better inform parents, through public information campaigns, of the tools that exist for them and their role in keeping children safe online.

Question 13: What requirements should the Code contain to ensure that VSPS provide for effective media literacy measures and tools?

As part of our Global Strategic Response, the [Education and Outreach Framework](#) asserts that children, parents and caregivers, and the public in general need education on safe and responsible digital use so that they are aware of the risks, know what is expected of them and can respond appropriately to negative situations or harmful or inappropriate content. The skills and competences that users need to be able to participate as responsible digital citizens are not acquired automatically and need to be learned, practised and provided for. Some core areas to cover in education content include:

- competent and positive engagement with digital technologies, e.g., digital literacy (inclusion, access, creating, learning, working, communicating, playing);
- active and responsible participation in global online communities (rights, responsibilities, ethics, health, values, attitudes, intercultural engagement, community engagement, e-presence, ways of communicating); and
- balancing digital and offline worlds (safety and risks, wellbeing, privacy, informal vs formal settings, consumer awareness, evaluating content).

Education and skills building should be delivered through accessible channels that are appropriate to age, gender, race, disability, culture, nationality and language. Both social and emotional learning concepts should also be included in online safety education to support children in developing their social and emotional skills to engage in respectful online relationships and strengthen resilience.

Question 14: How should we ask VSPS providers to address online harms in their terms and conditions in the Code, including the harms addressed under Article 28b? How should key aspects of terms and conditions be brought to users' attention? What examples are there of best practice in relation to terms and conditions including content moderation policies and guidelines?

The recent provision of Article 28b (added to the EU's Audiovisual Media Services Directive in 2020) requires platforms to take measures to address a range of online harms, including, child sexual abuse content, terrorist content, hate speech, incitement to violence, misleading advertising and the spread of misinformation. Regarding child sexual abuse content specifically, VSPS terms of services should clearly state that the platform has a zero-tolerance approach

regarding child sexual exploitation and abuse on its services. Terms and conditions should be concise and clear in defining what child sexual abuse material entails (photographs, videos, live streaming, grooming and digital or computer generated images, including the current emerging threat of AI-generated content), that such material is prohibited, how users can report CSAM and what the consequences will be for posting such content (ban from platform, referral to law enforcement, investigation and possible prosecution).

Question 15: How should we ask VSPS providers to address content moderation in the Code? Are there any current practices which you consider to be best practice? How should we address automated content detection and moderation in the Code?

Automated content detection and moderation are essential elements of the response to tackle child sexual exploitation and abuse online. There are different ways in which automated technologies can be used to detect, report, remove and block child sexual abuse online. Effective detection of 'known' child sexual abuse material is made possible by two linked techniques called 'hashing' and 'hash-matching'. These techniques have significantly accelerated the identification and removal of known child sexual abuse material from the internet. In addition to these techniques, the development of AI classifiers has been incredibly useful in the detection, reporting, removal and blocking of 'unknown' or 'new' child sexual abuse material online. Such automated or semi-automated moderation systems identify harmful content by following rules and interpreting many different examples of content which is and is not harmful. In a [2021 survey of tech company practices, conducted by WeProtect Global Alliance and the Tech Coalition](#), 84% of the companies surveyed said they had at least partly automated processes for forwarding reports of child sexual abuse online, suggesting that report management is relatively efficient. VSPS should continue to work – in partnership with safety tech experts and industry – on enhancing the accuracy of classifiers to detect 'unknown' child sexual abuse content (including livestreamed content) and grooming in both non-encrypted and encrypted video sharing environments. Open sourcing (with appropriate controls in place) should be used to encourage collaboration between relevant actors and help set consistent standards for safety technologies.

Question 16: What requirements should the Code include about procedures for complaint-handling and resolution, including out-of-court redress or alternative-dispute resolution processes? To what extent should these requirements align with similar requirements in the DSA? What current practices could be regarded as best practice? How frequently should VSPS providers be obliged to report to the Commission on their complaint handling systems and what should those reports contain? Should there be a maximum time-period for VSPS providers to handle user complaints and if so, what should that period be?

Requirements should be harmonised with the Digital Services Act. The DSA requires online platforms to have clear and transparent complaint-handling procedures, and to provide users with a fair and effective way to resolve their complaints. Article 17 covers the obligation for platforms to provide an internal complaint handling system and Article 18 obliges online platforms to engage with certified out-of-court dispute settlement bodies to resolve any dispute with users of their services.

Question 17: What approach do you think the Code should take to ensuring that the safety measures we ask VSPS providers to take are accessible to people with disabilities?

Our 2021 research, conducted by Economist Impact, titled [Estimates of childhood exposure to online sexual harms and their risk factors](#) found that young people who self-identified as disabled appear to be more vulnerable to online sexual harms than those who did not self-identify as disabled (57% v 48% experienced at least one online sexual harm). Much of this vulnerability was a result of being targeted by an adult they knew. Our briefing paper [The sexual exploitation and abuse of deaf and disabled children online](#), written in partnership with DeafKidz International and Childhood USA, states that children with disabilities should have full access to safety and protection programmes that allow them to stay safe online. The paper found that there is a significant gap in the data on the sexual exploitation and abuse of disabled children online which means that it is currently not possible to accurately know the level of incidence or prevalence. Specific and dedicated research that engages the wider disability community is therefore required before designing and implementing solutions. Video sharing service providers have a responsibility to develop and implement policies and procedures to protect children and adults living with disabilities from online harms. These policies and procedures should be tailored to the specific needs of children and adults living with disabilities. Video sharing service providers should be required to provide training to their moderators on how to identify and remove harmful content that is targeted at children and adults living with disabilities and design reporting and blocking tools in an accessible way. VSPS providers should conduct research into lived experiences and work with disability organisations to better understand accessibility issues and to consequently remedy them. It could also be useful to ask that VSPS providers collect data on the types of harmful content that is targeted at children and adults living with disabilities. This data could then be used to identify the specific risks that these groups face online and to develop more effective policies and procedures to protect at-risk groups.

Question 18: What approach do you think the Code should take to risk assessments and safety by design? Are there any examples you can point us towards which you consider to be best practice?

In recent years, risk-based regulatory approaches have increased in momentum and popularity. Regulations that require providers to assess risks posed to all children (not just child users – as harms can have an indirect broader impact) and to design and operate services in such a way as to mitigate specific risks, have most potential to curb trends and encourage Safety by Design by helping to prevent exploitation and abuse from happening in the first place. Risk assessments should serve to identify, analyse, and assess the systemic risks that VSPS pose to fundamental rights, the internal market, and public order. The EU’s Digital Services Act identifies some key systemic risks:

- The dissemination of illegal content;
- Negative effects for the exercise of fundamental rights;
- Negative effects on civic discourse and electoral processes, and public security; and
- Negative effects in relation to gender-based violence, the protection of public health and children and serious negative consequences to the person’s physical and mental well-being.

Under EU law, Very Large Online Platforms (VLOPs) and Very Large Online Search Engines (VLOSEs) are required to carry out risk assessments at least once a year. The risk assessments

must be specific to their services and proportionate to the systemic risks identified. The risk assessments must be specific to their services, proportionate to the systemic risks identified and consider:

- The nature of their services;
- The size of their user base;
- The geographic scope of their operations;
- The impact of their services on fundamental rights, the internal market, and public order.

Risk assessments are useful in encouraging platforms to be safe by design since they lead to platforms developing safety plans and implementing safety features.

Question 19: How do you think that cooperation with other regulators and bodies can help us to implement the Code for VSPS?

Given the transnational nature of the internet, international cooperation is an important part of the [Global Strategic Response](#) to tackle online harms, such as online child sexual exploitation and abuse. The national regulator should therefore have the powers to collaborate with other international regulators who are working to tackle online harms. Cooperation with other regulators in the form of sharing information and good practice, updates on new research and tools and identifying areas where regulators can collaborate to tackle cross-border harms will bolster the response and ensure that perpetrators of harm are held accountable despite their location. Concrete projects that regulators can embark on together include identifying common threats and developing harmonised responses and investing in capacity building projects to ensure that all countries have the means to tackle harm online. When it comes to international partnership, clear goals and objectives will have to be defined as well as a framework for cooperation.

Question 20: What approach do you think we should take in the Code to address feeds which cause harm because of the aggregate impact of the content they provide access to? Are there current practices which you consider to be best practice in this regard?

VSPS should take a risk-based approach in assessing feeds which cause harm because of the aggregate impact of the content they provide access to. By first assessing the aggregated risk and impact of feeds and channels and then implementing actions to remedy or limit the harm will be necessary. Possible interventions by the platforms could include prohibiting such content from featuring in autoplay functions and recommendation lists, to the suspension of accounts and content in more serious cases. While platforms should provide users with the necessary tools to control the content that appears on their feeds, they also have the responsibility to identify feeds/channels that repeatedly put users, including children, at risk.

Question 21: Do you have any views on how requirements for commercial content arranged by a VSPS provider itself should be reflected in the Code?

N/A

Question 22: What compliance monitoring and reporting arrangements should we include in the Code?

Transparency reports should be a common minimum standard for all VSPS. Video sharing platform services should be transparent about their policies and procedures for addressing harmful content. This includes providing information about how they identify and remove harmful content, how they respond to user reports, and how they measure the effectiveness of their policies and procedures. Clarifying which specific information is required for transparency reports and encouraging companies to be clear/detailed in their reports will be essential in ensuring that the most helpful and accurate information is being shared by the platforms.

Question 23: Should the Code have a transition period or transition periods for specific issues? Which areas touched on in this Call for Inputs may VSPS providers require time to transition the most? What time frame would be reasonable for a transition period?

Transition periods are important to allow industry enough time to adapt and comply with new rules and regulations, from developing new systems and processes for moderating content to establishing new relationships and processes with law enforcement authorities. These can be particularly complex issues to iron out and it is important to get the details right. If the code was to be implemented in a staggered approach, priority should be given to harms where the severity and impact are greatest, such as online child abuse and exploitation.

This submission is reflective of the views of the Secretariat of the WeProtect Global Alliance and does not necessarily represent the opinions and positions of any of its members.

For further information, please contact Eleanor Linsell, Policy & Advocacy Manager, at WeProtect Global Alliance:

████████████████████



Input Commissariaat voor de Media (Dutch NRA)

Developing Ireland's First Binding Online Safety Code for Video-Sharing Platform Services

September 2023, Hilversum

Introduction

In response to the Call for Inputs from the Coimisiún na Meán (the “Commission”) to inform a future consultation for the Online Safety Code (“the Code”) for Irish-based video sharing platforms (VSPs), we, the Commissariaat voor de Media (“Commissariaat”), are honoured to share our views. We would also like to thank the Commission once again for the hospitality shown to us during our visit to the Dublin office on the 22nd of March this year, which provided us with both the distinct honour of visiting you in person and the opportunity to speak to you about your regulatory plans for the future.

The Commissariaat recognises the importance of a strong code of conduct for Irish-based VSPs and thus welcomes this consultation. We hereby wish to thank the Commission for giving us the opportunity to respond. First and foremost, a strong Online Safety Code is important for ensuring the protection of all EU citizens who are active on these platforms.

However, as a national media regulator who is responsible for supervising the active users on these platforms (“the video uploaders”), we face our own challenges with respect to the interplay between these two applicable regulatory systems. These challenges arise particularly in relation to topics for which there is currently no European harmonisation, such as the protection of minors. Therefore, it is crucial that these rules are aligned and that any loopholes are addressed. Finally, Ireland will also serve as an example for other European regulators who will eventually be tasked with providing oversight over VSPs. In this respect, the Code will also provide strong inspiration for our own future regulation of VSPs.

We will respond to a selection of questions that we have taken from the Call for Inputs where we felt our input could be of most value. We hope that you will find our input useful and, of course, we are more than happy to clarify any questions you may have.

Selected questions

Question 1: What do you think our main priorities and objectives should be in the first binding Online Safety Code for VSPs? What are the main online harms you would like to see it address and why?

General mapping of online harms

The Rathenau Institute (a Dutch government-founded institute that was given the specific task of researching the impact of technology on our lives and society as a whole) published a research study in 2022 on Harmful Behaviour Online: An investigation of harmful and immoral behaviour online in the Netherlands¹:

¹ [Harmful Behaviour Online | Rathenau Instituut](#)



This study was the first to map out all the different elements of harmful and immoral online behaviour in the Netherlands. The Rathenau Institute developed a taxonomy of six categories of harmful and immoral conduct online, listing twenty-two different phenomena that all internet users in the Netherlands may encounter at some point (see Figure 1).

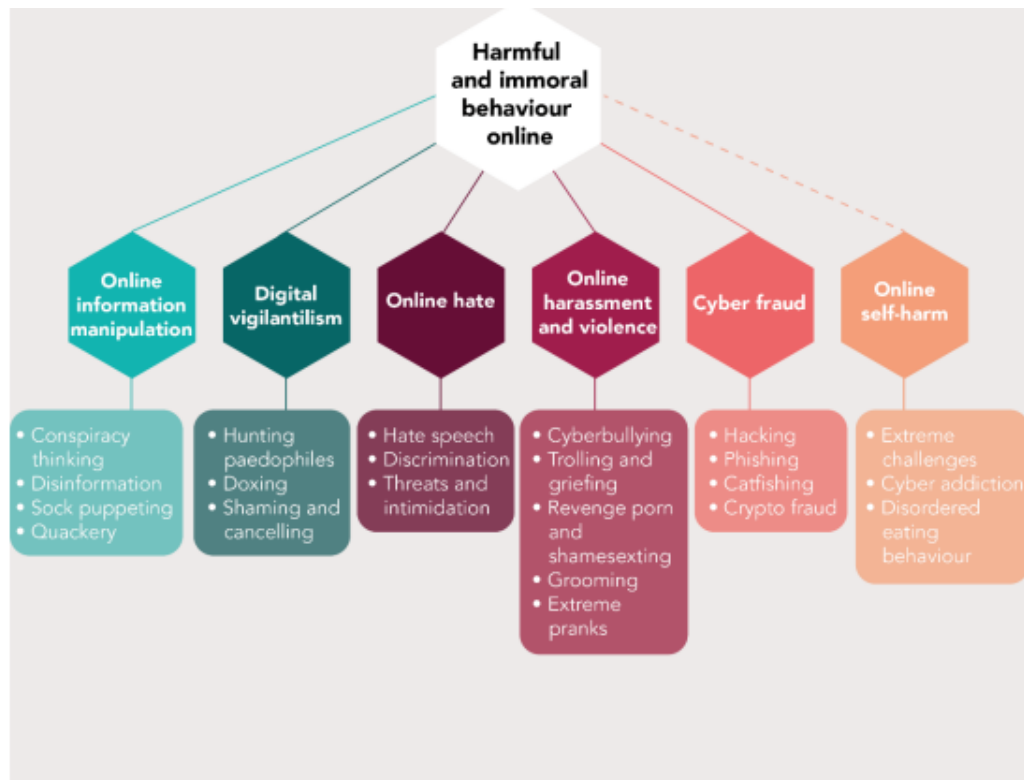


Figure 1 Taxonomy of harmful and immoral behaviour online². Source: Rathenau Instituut

The harmful behaviours listed in this taxonomy can severely impact upon individuals, groups, and society as a whole. These behaviours can range from a teenage girl starving herself in response to encountering extreme challenges with peers or discouraging female journalists and scientists from speaking out online in fear of online harassment, to societal disruption due to the spread of conspiracy theories and disinformation. Interviews with experts and the literature on the nature and scale of the phenomena listed in the taxonomy make it abundantly clear that all Dutch people are at risk of becoming involved in such behaviour, either as a victim, perpetrator, or bystander. Ultimately, as the report outlines, everyone can be affected by harmful and immoral behaviour.

The aforementioned research study emphasises that no distinction should be drawn between the identified online harms and the degree to which they can cause harm. The areas of online harm identified in the Call for Input (numbered 2 and 3) were also highlighted in this study.

One specific example of harmful content that should be considered pertains to those technologies that are used in videos to manipulate reality in any form, such as Augmented Reality (AR), or deepfakes in video content. Examples of this are the digital filters used by video uploaders/influencers to alter their appearance or situations in which deepfake-technology is



used.

Suggestion of an appropriate measure that can be taken to address this type of online harm: the VSPs should create a function to address when the video in question contains AR or deepfakes or any other sort of audiovisual manipulation.

Online harms that are relevant to the supervision of video uploaders

In the Netherlands, our supervision of video uploaders who are active on VSPs is focused on two main areas of online harm. In the Call for Inputs document (p. 7), these two areas are numbered as the first and fourth main areas of online harm addressed in Article 28b (2) of the AVMSD: 1) content that might impair the physical, mental, or moral development of minors, and; 4) commercial communications including advertising, sponsorship, and product placement, with a specific focus on commercial communications directed towards minors. These are the biggest and most common types of online harm, especially with respect to video uploaders.

Question 2: What types of online harm do you think should attract the most stringent risk mitigation measures by VSPs? How could we evaluate the impact of different types of harm, e.g., severity, speed at which harm may be caused? Is there a way of classifying harmful content that you consider it would be useful for us to use?

Out of the four main areas of harmful content outlined in the Call for Inputs document (p. 7), it is our belief that 2) content that incites violence or hatred against a group of persons or a member of a group based on any of the grounds referred to in Article 21 of the European Charter of Fundamental Rights and 3) Content, the dissemination of which constitutes a criminal offence under EU law, require the most stringent risk mitigation measures. This is based on the aforementioned research by the Rathenau Institute, which identified hate speech, discrimination, and threats of intimidation as some of the most severe risks online. Despite the fact that the research was conducted in the Netherlands, we nevertheless contend that the observations and findings are universal and most likely would not significantly differ to the situation in other European countries.

Of course, this does not mean that we see the other areas as being less important; rather, it is our contention that less stringent measures can be put in place for these types of harmful content, such as content that may impair the physical, mental, or moral development of minors.

With respect to the classification of harmful content (for minors), we would like to refer you to our national self-regulatory classification institute: NICAM², who established the Kijkwijzer classification system, PEGI and YouRateIt. They developed these classificatory instruments based on scientific research: *'You Rate It is a simple tool for consumers to classify their content on video sharing platforms such as YouTube. This way, children and teenagers can be warned about the imagery. You Rate It was developed by NICAM for use on an international level.'*³

Question 3: Do you have reports, academic studies or other relevant independent research that would support your views? If you do, please share them with us with links to relevant reports, studies or research.

² [Our mission | Nicam](#)

³ [You Rate It | Nicam](#)



- NICAM carries out extensive research on the behaviour of minors on platforms: [Research | Nicam](#)
 - In 2022, the Rathenau Institute (a Dutch government-founded institute that was given the specific task of researching the impact of technology on our lives and society as a whole) published research on Harmful Behaviour Online: An investigation of harmful and immoral behaviour online in the Netherlands: [Harmful Behaviour Online | Rathenau Instituut](#)
- There is an English Summary available here: [Titel \(rathenau.nl\)](#)

Question 4: What approach do you think we should take to the level of detail in the Code? What role could non-binding guidance play in supplementing the Code?

We would recommend opting for option 3 - *A mixed approach*.

Ideally, there would be a combination of rules and principle-based approaches, so that VSPs are encouraged to come up with their own solutions to achieve the prescribed outcomes. Content that is seen or classified as being more harmful, such as criminal offences and hate speech, could have more specific rules than content deemed to be less damaging, such as commercial influence. In the Call for Input document the Commission states: *'We could also require VSPS providers to be transparent about the measures they are taking to comply with high-level requirements and to provide metrics that would enable their effectiveness to be assessed.'* (p. 10). We wholeheartedly support this idea on the grounds that we believe transparency over the enforcement and effectiveness of these measures is of paramount importance for both the further development of VSPs regulation in the EU and the evaluation of this specific Code in the future.

Question 5: What do you think would be the most effective structure for the Code? What are the most important factors we should consider when we decide how to structure the Code?

We would recommend choosing one of the two first options, rather than solely basing the structure of the Code on Article 28b (3) of the AVMSD, insofar as this might make it harder to align the Code with the DSA.

Question 7: To what extent, if at all, should the Code require VSPS providers to take measures to address content connected to video content?

We believe it would also be a good idea to include measures in the Code to address content that either accompanies or is linked in other ways to the video content. This is because it is not only audiovisual content itself but also the descriptions under the videos that can be harmful and/or influence how users interpret the video.

The comments section is an important part of social media platforms, insofar as it allows video uploaders to interact with their subscribers and fans. Media are becoming more interactive and cross-medial in general. There is often an entire community (often with websites and sometimes even events) behind video uploaders' accounts, and, hence, it is important to also consider this when regulating VSPs as opposed to only focusing on the audiovisual content itself.

Flagging mechanisms should also be implemented for the comments section, so that the discussions that take place there will also meet the standards of the AVMSD. More stringent measures should be taken towards hate speech and threats towards video uploaders, journalists and/or marginalised groups.

Question 8: How should we ask VSPS providers to introduce a feature that allows users to declare when videos contain advertising or other type of commercial communications? Should



the Code include specific requirements about the form in which the declaration should take? What current examples are there that you regard as best practice.

It is important that the way of declaring commercial communications is easily visible. During our supervision, we encounter cases in which our supervisory team are unable to find the declaration, either because it is so small or because it is in a distinctive colour. In some instances, the format contains a white font which is not sufficiently visible in a video with a white background. VSPs should thus make sure that there are multiple options in terms of colour/background for the declaration.

Another issue we have encountered in practice is that after accepting the cookie policy whilst watching the video, the declaration may not show up when watching the video again. This should not be allowed.

Furthermore, we have noticed that video uploaders often use multiple hashtags with only the final one containing the declaration of, for example, advertising. This declaration should be the first hashtag. It would also be a good idea to urge VSPs to constantly test what works best. This could be done with A/B testing, for example.

Transparency is vitally important, particularly when it pertains to the location and contact details of the service providers who are active on VSPs. Under the Dutch Media Act, the registered video uploaders should state that they are registered with The Commissariaat (Dutch NRA) and disclose their contact information for any complaints. Uploaders on VSPs should have an “About me” page. Whilst on YouTube, for example, there is enough space for uploaders to elaborate, but on TikTok there is limited space (under 100 characters). When there is limited space, as is the case on TikTok and Instagram, it becomes harder to mention this on their page. The result is that users are not aware that these uploaders must abide by the Dutch Media Act. The “About me” page is also important in terms of providing transparency over the identity of the uploader on VSPs.

Uploaders do not always provide their true country of residence to VSPs, which means that on the uploader homepage the real country from which the uploader operates is not visible. If it was mandatory for the uploader to state their country of residence or operation, then this would make it easier for media regulators in other European countries to assess if an uploader needed to register in their country, which, in turn, would increase the level of transparency about their identity.

Question 9: How should we ask VSPS providers to introduce and design a flagging mechanism in the Code? How can we ensure that VSPS providers introduce the mechanism in a user-friendly and transparent way? How should we ask VSP Providers to report the decisions they've made on content after it has been flagged? To what extent should we align the Code with similar provisions on flagging in the DSA?

It would be in users' best interests to make the flagging of content as easy as possible. Users are in most cases not going to be aware of whether their complaint is based on the DSA or the Code (AVMSD). It is therefore the responsibility of the platform to determine this behind the scenes. The flagging mechanism should provide users with a list of different reasons, and therefore types of harmful content, so that the VSP can decide both what types of measures have to be taken and which regulation this is based on. This should then be reported back to users. However, VSPs



should also be transparent about the rules that apply in their case and whether their flagging is going to be processed or not. In the event that VSPs decide, based on their Terms & References, not to process the complaint, then this decision should be clearly explained to users.⁴

In our opinion, there should be a function for regulators to flag/file a complaint and VSPs should prioritise flagging deriving from regulatory authorities. For example, when a regulatory authority flags certain content which does not comply with their local rules and regulations, then VSPs should respond to these flags in a timely manner.

Question 10-I: What requirements should the Code include about age verification and age assurance? What evidence is there about the effectiveness of age estimation techniques? What current practices do you regard as best practice?

Firstly, we think that self-declaration is **not** an appropriate age-verification tool, insofar as it does not actually verify someone's age and is easily worked around. The most robust age verification tools are often based on biometric data and provided by third parties. A good example of such a system, which has also been approved by the German Kommission für Jugendmedienschutz (KJM)⁵, is an age verification system like Yoti.⁶ We have discussed this topic along with best-practice examples in the EPRA AI Taskforce during a session on how AI applications like Yoti⁷ can protect minors online.

Question 10-II: What sort of content should be shown by default to users who are logged out or in private browsing mode and whose age cannot be verified or assured?

This should only be content that the VSPs can determine with certainty as not being harmful to the physical, mental, and moral development of minors.

A good example of how to set this up is the way that Ofcom did in their VSP regulation⁸, which is to divide VSPs into segments/risk groups (let's say A, B and C) and depending on the segment, age verification measures should then be taken. Risk factors could include adult content such as pornography, risk of harmful videos going viral on the VSP, type of audience.

Question 11: What requirements should the Code have in relation to content rating? What do you consider to be current best practice? What experiences have you had using content rating systems on platforms and do you think they have been effective? What steps could we ask VSPs to take to ensure content is rated accurately by users?

In terms of age rating and classification systems we would like to advocate for a European classification system which applies to all European countries and takes into account cultural differences. We also discussed this during our visit to your Dublin office on the 22nd of March this year together with our Ministry for Education, Culture and Science and NICAM. The Code represents an excellent opportunity to take a first step in this direction to harmonise age ratings and content classification in Europe, since VSPs have users spread across the EU.

⁴ In line with article 14 DSA.

⁵ [KJM bewertet sieben weitere Altersverifikationssysteme positiv - Pressemitteilungen - KJM \(kjm-online.de\)](#)

⁶ [Age verification tools for online customers and custom-built apps - Yoti](#)

⁷ Summary EPRA 3rd AI Roundtable: [AI_Roundtable_3_summary.pdf \(epra.org\)](#)

⁸ [Video-sharing platform guidance \(ofcom.org.uk\)](#)



Currently, video uploaders in the Netherlands are required to use a content classification system that is similar to the one used to classify and rate content for broadcasters and VODs. The advantage of this is that users are familiar with the symbols. According to research by NICAM⁹, the Kijkwijzer-system is considered to be valuable by most European parents who also indicated that they understand the system. Given that content uploaded on VSPs can be viewed throughout Europe, it is important that minors and non-English speakers are able to understand the ratings and make informed decisions based on the content ratings.¹⁰

However, some platforms only use warnings such as 'contains sensitive content', which minors in particular find to be overly vague, not to mention that it is unclear from whom the warning originates. Moreover, the notification also does not stand out and is sometimes incorrect. It is also not possible for the uploader to assess the videos in advance and assign warnings to them. Finally, most video uploaders are not located in the Netherlands or Europe, which means that most video uploaders who upload harmful content, are currently not obligated to use a content rating system.

We recommend that content classifications should be "fed" into the algorithm used by the platform, so that young users are not exposed to harmful content. All users should be able to filter certain harmful content. The platform Twitch, for example, currently requires users' explicit consent before each video that contains a Content Classification Label. The Content Classification Label has specific categories, such as Gambling and Violent and Graphic Depictions, that makes it clear to users what content they are consenting to see.

Finally, all platforms should implement a system that makes it easy for uploaders to rate harmful content prior to uploading. The solution for this would be to embed the use of age ratings and content pictograms within the VSP. Hereby allowing uploaders to show the age and content ratings on the platform next to the title of a production as well as embedding them during the first five seconds in their video on a 'ratings layer'. We recommend to include the obligation in the Code to facilitate (national) rating systems on their platforms by providing their uploaders with options to embed and show ratings in their videos.

Question 12: What requirements should the Code have in relation to parental control features? How can we ensure that VSPs providers introduce the mechanism in a user-friendly and transparent way? Can you point to any existing example of best practice in this area? Should parental controls be 'turned-on' by default for accounts of minors or where age is not verified? Can you point to any existing examples of best practice in this area?

From our experience with international commercial video on-demand services who are based in the Netherlands, such as Disney+¹¹ and Netflix¹², it is evident that these parties have extensive experience with parental control measures. For example, both Disney+ and Netflix have the option to set content ratings for each profile. This allows for flexibility over what is appropriate for each user, and the content rating can be adjusted as a minor gets older and other content becomes appropriate.

⁹ Yearly report is only available in Dutch: [NICAM Jaarverslag 2022](#).

¹⁰ And as already mentioned in our answer to Question 2: NICAM also developed YouRateIt, which is specifically designed for users on platforms: [You Rate It | Nicam](#)

¹¹ [Parental Controls on Disney+ | Disney+ \(disneyplus.com\)](#)

¹² [Parental controls on Netflix](#)



Should parental controls be 'turned-on' by default for accounts of minors or where age is not verified?

Turning on parental controls by default prevents minors from attempting to find a way around age verification or the parental controls set by their parents.

Question 13: What requirements should the Code contain to ensure that VSPs provide for effective media literacy measures and tools?

We also see notable opportunities for the Code to promote media literacy. Therefore, it would be useful for the Code to pay attention to those measures pertaining to Media Literacy that are stated in Article 28b (3) (f) of the AVMSD.

Since the promotion of media literacy does not fall under our legal mandate, our own practical experiences are limited. Nevertheless, we would like to refer to an ERGA report from 2021: ERGA Media Literacy Report Recommendations for key principles, best practices, and a Media Literacy Toolbox for Video-sharing Platforms.¹³ Amongst other things, this report outlines how six key principles of media literacy initiatives can be implemented by VSPs. Given our lack of experience in the promotion of media literacy on VSPs, we would once again refer to the best practice in this field from Ofcom.¹⁴

Question 19: How do you think that cooperation with other regulators and bodies can help us to implement the Code for VSPs?

First of all, cooperation on an international level is vitally important not only for cross-border cases but also in terms of overlapping supervision. Currently, many European regulators are setting up their supervision of video uploaders who upload their content on these VSPs.

In this process, cooperation is fundamentally important. The national rules in EU countries that apply to video uploaders are largely based on the VOD rules from the AVMS Directive and are as yet not fully harmonised. Within ERGA, we try to ensure that the implementation of the AVMS Directive, including within the area of the regulation of video uploaders (also called 'vloggers' in ERGA Subgroups) and supervision, is as consistent as possible.¹⁵ However, those Member States who have already set up their supervision of video uploaders also face many practical challenges in terms of both the supervision and enforcement of these rules. We are currently indexing these practical challenges within ERGA Subgroup 1 and a report providing guidance on how to achieve greater consistency and uniformity in national approaches will be published towards the end of this year.

It would also be beneficial to establish a European VSP working group in which all the regulators who are faced with the supervision of VSPs could cooperate, and share their challenges and best practices. The VSP Regulation group within EPRA, which provides workshops, is a good example of such a group.

¹³ [ERGA Media Literacy Report Recommendations for key principles, best practices, and a Media Literacy Toolbox for Video-sharing Platforms](#)

¹⁴ [Video-sharing platform guidance \(ofcom.org.uk\)](#), Section 4 of their Guidance document, pp. 49-53.

¹⁵ In the last two years ERGA published two reports on the regulation of vloggers to contribute to this goal of consistent implementation of the AVMS Directive: Report – [How to identify and localise vloggers and regulate their commercial communication?](#) (2022); Report – [Analysis and recommendations concerning the regulation of vloggers](#) (2021).



Question 20: What approach do you think we should take in the Code to address feeds which cause harm because of the aggregate impact of the content they provide access to? Are there current practices which you consider to be best practice in this regard?

The non-profit Centre for Countering Digital Hate (CCDH) published a paper, 'Deadly by Design'¹⁶, that analyses harmful content on TikTok and puts forward several recommendations. The researchers conducted extensive research on the recommendation algorithm used by TikTok to provide users with a 'For-You' feed. They found harmful content related to eating disorders and self-harm and suicide was being recommended and going viral. They also discovered that TikTok regularly and purposefully recommends harmful content to vulnerable minors.

In their paper, the CCDH calls for global standards to reform social media, based on the fact that the platforms have a global reach, and recommends a framework through which to achieve this. First, they suggest safety by design, which includes amending products and services to embed safety considerations. Second, they recommend prioritising transparency over the algorithms and rule enforcement. Third, accountability should be improved by allowing independent enforcement and the possibility to challenge decisions and omissions. Finally, companies and senior-level executives should be held responsible for implementing safety considerations as well as the consequences for actions or omissions that lead to harm. This proposed framework by the CCDH is further explained and substantiated in their paper 'A Global Standard for Regulating Social Media'¹⁷.

Question 21: Do you have any views on how requirements for commercial content arranged by a VSPS provider itself should be reflected in the Code?

The commercial content offered by VSPs should also be in compliance with general advertisement rules from the AVMS Directive. In particular, rules for commercial content targeting minors should have a special place in the Code. Based on the Dutch Advertisement Code, there are specific, stricter rules in place for advertisements targeting minors. Commercials targeting minors should neither be misleading in any sense nor cause them moral or physical harm. It goes without saying that minors are a more vulnerable audience and, as such, easier to mislead.¹⁸

The Code should thus include general advertising rules that VSPs must comply with as well as specific and stricter rules on advertising targeting minors and children's accounts. VSPs should guarantee that they can identify which users are minors and those who are not, so that they can be sure that minors will only be exposed to advertising that meets the strictest requirements. Hence, in the event that the VSP has not yet been able to identify whether or not a user is a minor, then the advertising will have to meet the most stringent requirements for advertising offered to that user.

Furthermore, all these measures must also take into account the following rule from Article 28ter (3) of the AVMSD: minors should be protected against inappropriate advertisement without

¹⁶ [CCDH-Deadly-by-Design_120922.pdf \(counterhate.com\)](#)

¹⁷ [Copy of STAR Framework for website \(counterhate.com\)](#)

¹⁸ [About the Stichting Reclame Code - Stichting Reclame Code](#)



collecting their personal data for commercial purposes such as direct marketing, profiling, and behaviourally targeted advertising.

As with all other advertisements, the commercial content arranged by VSPs should be required to be transparent. Both the advertisements and the advertiser should be clearly labelled, so that users are aware what content they are watching. In accordance with the DSA requirements, users should also be informed about why users are exposed to certain types of advertising. The Code should clearly require platforms to implement mechanisms that provide transparency to their users regarding all these advertising-related matters.

Question 23: Should the Code have a transition period or transition periods for specific issues? Which areas touched on in this Call for Inputs may VSPS providers require time to transition the most? What time frame would be reasonable for a transition period?

Whilst a transition period is certainly reasonable, it is important to note that VSPS providers are aware of the upcoming Code and, as such, are likely also aware of the likely content of the Code. Considering this fact and the importance of the subject at hand, a lengthy transition period thus seems unnecessary and undesirable. Especially those sections of the Code that deal with the most harmful content should have the shortest transition period, on the grounds that this type of content was already illegal prior to the Code entering into force, and, hence, these measures should already be in place.

A binding Code against toxic algorithms

ICCL submission to the Media Commission call for input on video-sharing platform services

SEPTEMBER 2023

Dr Johnny Ryan FRHistS
Senior Fellow, ICCL

In this submission

About ICCL.....	3
Summary: act on algorithms.....	4
Recommender systems	5
Prescriptive and verifiable	8
Action on algorithms	9
Notes.....	13

Contact: 

About ICCL

The Irish Council for Civil Liberties (ICCL) is Ireland's oldest independent human rights body. It has been at the forefront of every major rights advance in Irish society for almost half a century. ICCL helped legalise homosexuality, divorce, and contraception. We drove police reform, defending suspects' rights during dark times. In recent years, we led successful campaigns for marriage equality and reproductive rights.

Dr Johnny Ryan FRHistS is a Senior Fellow at ICCL. Previously he served in senior roles in technology and media. He is regularly invited to give expert testimony and has appeared before the European institutions and the U.S. Senate. His expert commentary has appeared in *The Economist*, *NATO Review*, and *The New York Times*.

Thanks to **Olga Cronin** and **Katarzyna Szymielewicz**.

Summary: act on algorithms

This submission demonstrates the hazard of platforms' algorithmic recommender systems, and proposes verifiable measures.

Selected Media Commission questions:

- Question 1 – “What do you think our main priorities and objectives should be in the first binding Online Safety Code for VSPS? What are the main online harms you would like to see it address and why?”
- Question 4 – “What approach do you think we should take to the level of detail in the Code? What role could non-binding guidance play in supplementing the Code?”
- Question 20 – “What approach do you think we should take in the Code to address feeds which cause harm because of the aggregate impact of the content they provide access to? Are there current practices which you consider to be best practice in this regard?”

Summary:

- Our submission focuses on **digital platforms' algorithmic amplification of hazardous content such as incitement to hate, violence and terrorism, racism and xenophobia**.
- We respond to questions 1, 4, and 20 of the Media Commission's invitation. Our answer to question 1 is the section “Recommender systems”; question 4 is the section “Prescriptive and verifiable”; and question 20 is the section “Action on algorithms”.
- The section “Recommender systems” shows that **platforms' recommender systems are particularly dangerous**. The section “Prescriptive and verifiable” shows that **platforms' voluntary and discretionary measures are ineffective**.
- We suggest several measures. Primary among them is that the Code should mandate that algorithmic recommender systems are not activated by default by platforms. **Toxic algorithms must stay off until a user decides to switch them on**. People must be able to use digital platforms without algorithms injecting poison into their feeds.
- Acting against algorithmic amplification rather than attempting to identify and unpublish harmful content is likely to be more effective, and **avoids intrusion upon the right to freedom of expression**.

Recommender systems

RESPONSE TO MEDIA COMMISSION QUESTION 1

Recommender systems are understood to be dangerous, and require prioritisation.

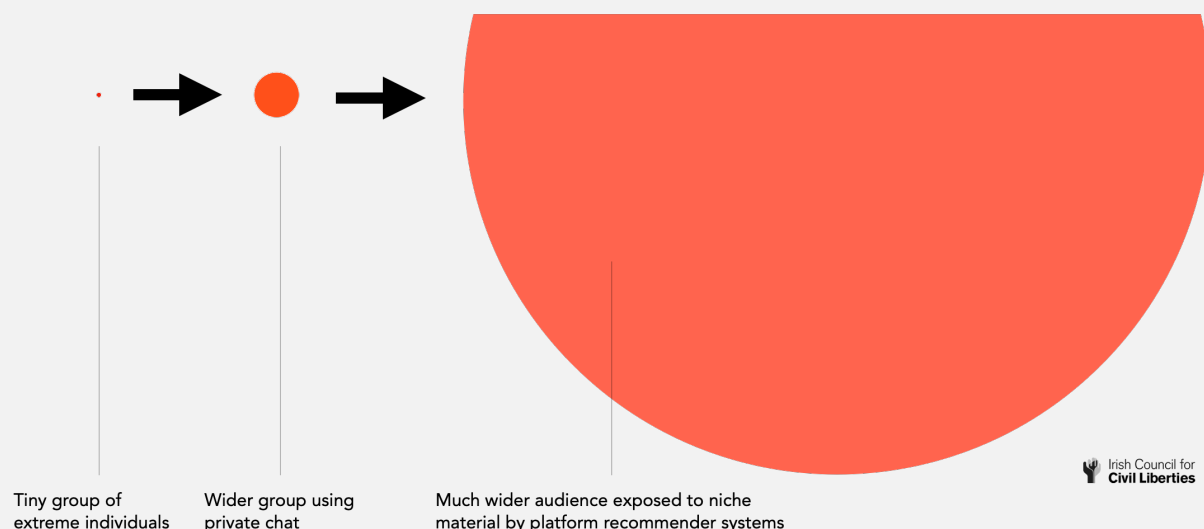
Examples:

- In August 2023 an Anti Defamation League study found that **Facebook, Instagram, and X (Twitter)** recommended **antisemitic and conspiracy content** to test users, including to users as young as **14 years old**.¹
- A global study of 37,000+ YouTube volunteers in 2022 showed that **most (71%) of the problematic² content they saw on YouTube was presented to them by YouTube's recommender system**.³ This new research followed YouTube recommender scandals and purported fixes by the company in preceding years.⁴
- In 2016 internal **Meta** research (later disclosed by whistleblower Frances Haugen) concluded that:

"64% of all extremist group joins are due to our recommendation tools... Our recommendation systems grow the problem".⁵ The researchers concluded: **"Our algorithms exploit the human brain's attraction to divisiveness."**⁶

Amplification of hate and hysteria

Digital platform **recommender systems** find emotive videos and posts and expose them to large audiences to maximise engagement. **Without algorithmic amplification, dangerous material from the small core group would not be widely seen.**



- An internal Meta document dated 2019 discussed “hate speech, divisive political speech, and misinformation” and noted:

“compelling evidence that our core product mechanics, such as virality, **recommendations, and optimizing for engagement, are a significant part of why these types of speech flourish on the platform. ... The mechanics of our platform are not neutral**”.⁷

- Another 2019 internal Meta document concluded that content moderation is impossible at large scale, and the focus should be on avoiding algorithmic amplification of the content:

“**We are never going to remove everything harmful** from a communications medium used by so many, **but we can at least ... stop magnifying harmful content** by giving it unnatural distribution”.⁸

- **United Nations investigators reported that Meta (Facebook) had played a “determining role” in Myanmar’s 2017 genocide.**⁹ Amnesty International’s follow-on investigation reported that Meta’s **algorithms** were essential contributors. Amnesty concluded that “**content-based solutions will never be sufficient to prevent and mitigate algorithmic harms**”.¹⁰
- The Irish Government’s National Counter Disinformation Strategy scoping paper noted in September 2023:

“New digital media and platforms can help to spread disinformation more quickly than ever before. Measures to counter this should enforce and incentivise the lawful use of people’s data, ethical business models, and **prevent digital platforms’ recommender algorithms from amplifying hate and hysteria in people’s video and social feeds for commercial gain**”.¹¹

- The **European Commission** reports that **Russian disinformation about its invasion of Ukraine** “was achieved through a combination of direct action by pro-Kremlin actors and **through algorithmic recommendation by the platforms**”.¹²

Recommendation:

- Recommender systems find emotive content and expose it to large audiences to maximise engagement. **Without this algorithmic amplification, dangerous material from a tiny number of extremists would not be widely seen.**
- As the examples above show, the content covered by section 139K(2)(c) OSMR is far broader than the illustrative examples in point 5.3.5 of the Media Commission’s request for input on recommender systems. **Since at least as early as 2016, digital platforms have understood that their recommender systems amplify hate and hysteria.**

- The Media Commission should therefore **prioritise acting against hazardous recommender systems** over other actions to tackle incitement to hate and violence, racism and xenophobia, and incitement to terrorism.*
- Acting against algorithmic amplification rather than attempting to identify and unpublish harmful content is likely to be more effective, and **avoids intrusion upon the right to freedom of expression**.

* This recommendation does not relate to harms such as bullying, self-harm, child sexual abuse, etc. Other measures, such as content moderation and tackling addictive design will be required for other harms.

Prescriptive and verifiable

RESPONSE TO MEDIA COMMISSION QUESTION 4

Voluntary and discretionary measures by platforms will not be sufficient.

Key insights:

- **Digital platforms have a very poor record of self-improvement and responsible behaviour**, even when lives are at stake as in Myanmar's genocide.
- Even when a platform understands the harm its recommender system causes, it is unlikely to voluntarily act. Despite internal concern about amplifying hazardous content, from 2017 to 2020 Meta strongly amplified¹³ posts that received "emoji" reactions from other people. Then, despite internal research in 2019 confirming that content receiving "angry emojis" was more likely to be misinformation, it persisted in strongly amplifying them until late 2020.¹⁴
- Digital platforms' voluntary measures against the risk they create are inadequate. In August 2023, the **European Commission** reported that voluntary measures taken by **YouTube, Facebook, Instagram, TikTok, Twitter, and Telegram** against Russian disinformation on their platforms had "failed".¹⁵ It concluded that **"Article 35 [DSA] standards of effective risk mitigation were not met in the case of Kremlin disinformation campaigns"**.

Recommendation:

- The Code must be **binding**. It must be robustly enforced, if necessary, by application for a blocking order to the High Court.
- Measures required by the Code **must be practical to monitor**. Our recommendations in response to question 20 are designed with this in mind.
- Digital platforms should have no opportunity to evade their responsibilities. **Clarity is essential** in the Code's specification of mandatory measures.

Action on algorithms

RESPONSE TO MEDIA COMMISSION QUESTION 20

Algorithmic recommender systems are optional - and highly hazardous - features rather than intrinsic elements of digital platforms.

Key insights:

- Section 139K(4)(a) OSMR provides that a Code may provide for “standards that services must meet, practices that service providers must follow, or **measures that service providers must take**”. The Media Commission is empowered to enforce those standards, including by way of an application to the High Court for a “blocking order” under section 139ZZC OSMR.
- Algorithmic recommender systems are neither legally nor technically essential components of digital platforms. The **European Court of Justice (CJEU)** ruled in July 2023 in *Bundeskartellamt v Meta* (including Facebook and Instagram) that **personalisation of content is “not objectively indispensable”**.¹⁶ In addition, platforms are required by Article 38 DSA to provide **alternative recommendations not based on a profile of the user**.
- Switching algorithmic recommender systems off is technically trivial. Virtually all websites and news media operate without such systems, **relying instead on the curatorial art of their editors**.
- There are alternative methods to curate a digital platform and show users a mix of memes, cat videos, celebrity news, and unboxing videos that do not require recommender systems which process profiles of each user. For example, platforms may rely on the user’s selection from a menu of the categories of content they are interested in, and have expert editors curate those categories of video and video creators.
- Digital platforms are required by **Article 9 GDPR** to have the person’s “explicit consent” to process “special category” personal data, including inferences about the platform user’s political views, sexuality, religion, ethnicity, health. These data cannot be processed for a recommender system unless the person has given their consent. **Any recommender systems that engage with a user’s politics, sexuality, religion, ethnicity, or health must be off by default**.

Recommendations:

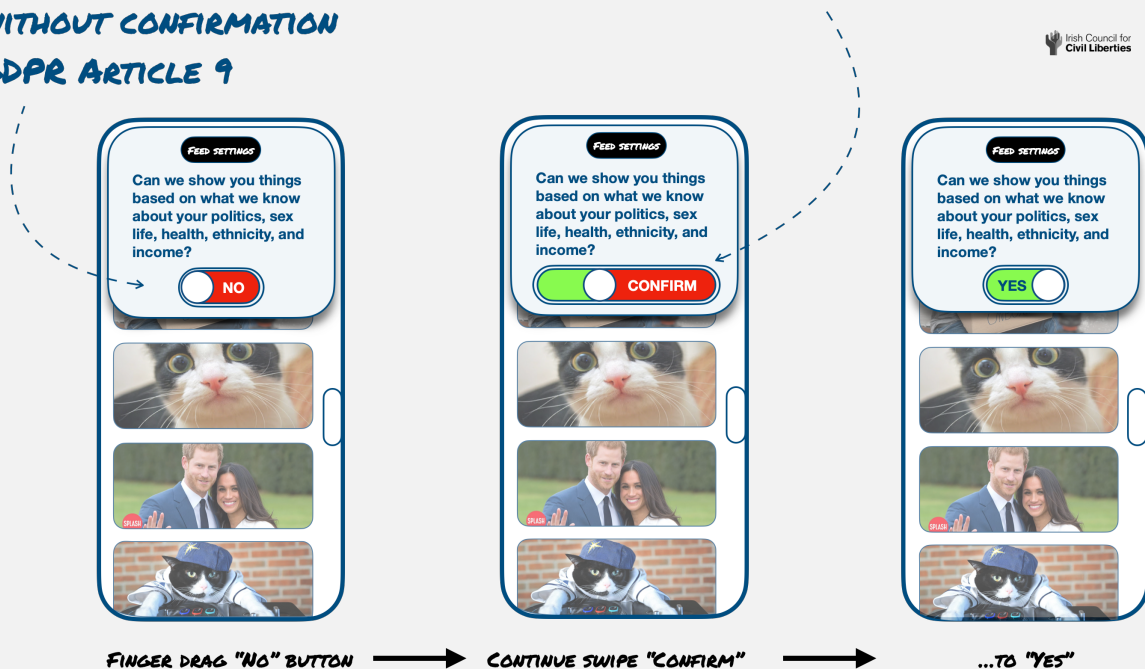
- The Code should mandate that algorithmic recommender systems are **not activated by default** by platforms. Users must be able to use a platform without being exposed to toxic algorithms that inject poison into their feeds.
- This should apply generally, but in particular to recommender systems that process (including by inference or proxy) “special category” data as defined by Article 9 GDPR. The GDPR prohibits processing of data about people’s **health, sexuality, political and philosophical views, religious beliefs and ethnicity**. The only applicable derogation for a platform is if a user has given “explicit consent”.
- The Code should require platforms to implement **lawful requests for explicit consent**.

Politics, sexuality, health... off by default

“Explicit consent” is understood to require a two-step action to give the person the opportunity to confirm their consent.¹⁷ Our indicative design two-step action is below.

**RECOMMENDER SYSTEM CANNOT PROCESS DATA ON USER'S SEX, HEALTH, POLITICS, OR RELIGION WITHOUT CONFIRMATION
GDPR ARTICLE 9**

**TWO STEP "EXPLICIT CONSENT" CONFIRMATION
GDPR ARTICLE 9 (2)(A)**



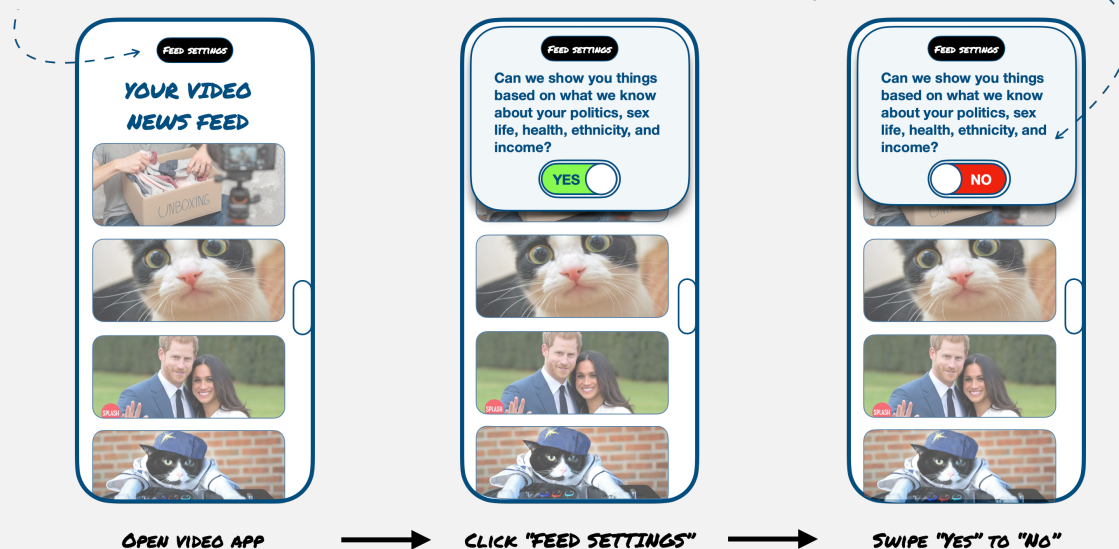
- The Code should require that **if a user activates a recommender system, then an immediately visible means of deactivating that recommendation system is shown prominently on the screen at all times** where the system is active, as provided for in DSA Article 27(1) and Article 38 of the DSA.

The DSA recommender system "off" switch

The Digital Services Act requires digital platforms to provide a recommender system off-switch, which must be visible at all times when the recommender system is active. Our indicative design for this is below.

**OPTIONS MUST BE VISIBLE WHERE THE RECOMMENDER SYSTEM IS ACTIVE
DSA ARTICLE 27(1)**

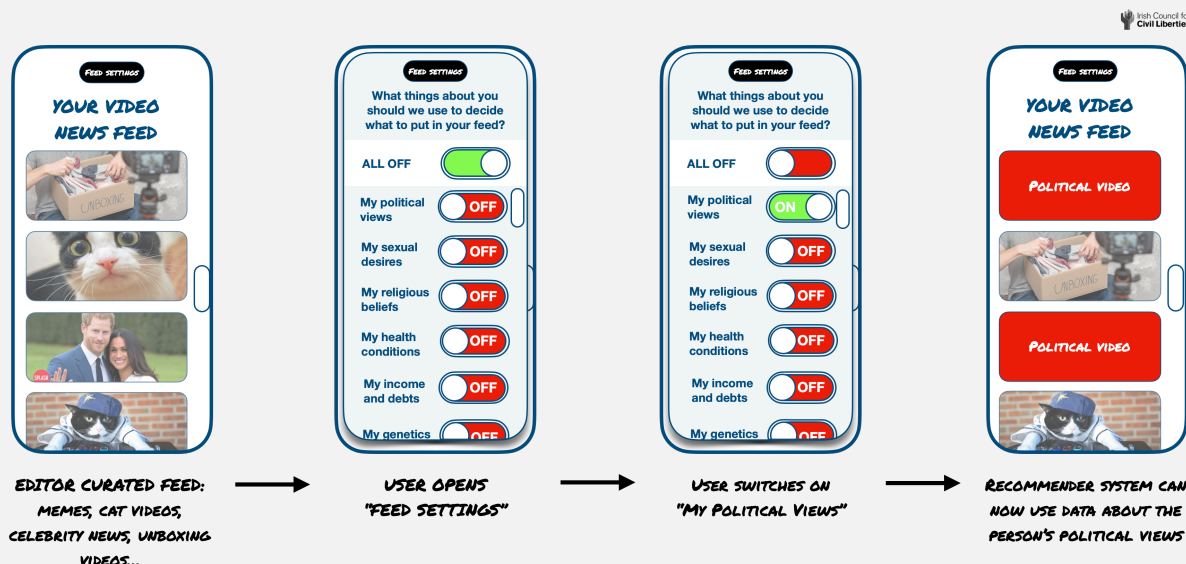
**MANDATORY OPTION FOR A RECOMMENDER SYSTEM NOT BASED ON A PROFILE
DSA ARTICLE 38**



- The Media Commission may wish to consider whether the Code should also mandate granular user control over the activation of recommender systems, including the types of data about the user available to a recommender system.

Granular control

A user may wish to receive algorithmic recommendations related to their financial situation without the recommender system also making inferences about other intimate aspects of their character and circumstances. Our indicative design for granular control is below.



- The Media Commission should be prepared for the possibility that platforms will respond with **"malicious compliance"**: implementing the least attractive designs and experiences for users in order to provoke outcry against regulatory intervention. For example, an entirely unedited and unordered feed of randomised video. However, digital platforms who maliciously comply create the risk that their users will depart to competitors who offer better service. Malicious compliance may be commercially damaging.

Notes

- ¹ "From Bad To Worse: Amplification and Auto-Generation of Hate", ADL, 16 August 2023 (URL: <https://www.adl.org/resources/report/bad-worse-amplification-and-auto-generation-hate>)
- ² "YouTube Regrets: A crowdsourced investigation into YouTube's recommendation algorithm", Mozilla, July 2021 (URL: https://assets.mofoprod.net/network/documents/Mozilla_YouTube_Regrets_Report.pdf), pp 9-13.
- ³ *ibid.* p. 17.
- ⁴ For example, see <https://www.theguardian.com/technology/2018/feb/02/how-youtubes-algorithm-distorts-truth> and YouTube's commitment to improve in 2019 <https://blog.youtube/news-and-events/continuing-our-work-to-improve/>.
- ⁵ "Facebook Executives Shut Down Efforts to Make the Site Less Divisive", Wall St. Journal, 26 May 2020 (URL: <https://www.wsj.com/articles/facebook-knows-it-encourages-division-top-executives-nixed-solutions-11590507499>).
- ⁶ "Facebook Executives Shut Down Efforts to Make the Site Less Divisive", Wall St. Journal, 26 May 2020 (URL: <https://www.wsj.com/articles/facebook-knows-it-encourages-division-top-executives-nixed-solutions-11590507499>); see also The social atrocity: Meta and the right to remedy for the Rohingya", Amnesty International, 2022 (URL: <https://www.amnesty.org/en/documents/ASA16/5933/2022/en/>), p. 54.
- ⁷ The Facebook Papers, "What is Collateral Damage?", 12 August 2019, p. 34, cited in "Internal Facebook documents highlight its moderation and misinformation issues", TechCrunch, 25 October 2021 (URL: <https://techcrunch.com/2021/10/25/internal-facebook-documents-highlight-its-moderation-and-misinformation-issues/>)
- ⁸ The Facebook Papers, "We are Responsible for Viral Content", 11 December 2019, p.17
- ⁹ U.N. investigators cite Facebook role in Myanmar crisis, Reuters, 12 March 2018 (URL: <https://www.reuters.com/article/us-myanmar-rohingya-facebook/u-n-investigators-cite-facebook-role-in-myanmar-crisis-idUSKCN1GO2PN>).
- ¹⁰ "The social atrocity: Meta and the right to remedy for the Rohingya", Amnesty International, 2022 (URL: <https://www.amnesty.org/en/documents/ASA16/5933/2022/en/>), pp. 45-48, p. 71.
- ¹¹ "Scoping Paper: National Counter Disinformation Strategy Working Group", Government of Ireland Department of Tourism, Culture, Arts, Gaeltacht, Sport and Media, August 2023, p. 11.
- ¹² "Digital Services Act: Application of the Risk Management Framework to Russian disinformation campaigns", European Commission, 30 August 2023 (URL: <https://op.europa.eu/en/publication-detail/-/publication/c1d645d0-42f5-11ee-a8b8-01aa75ed71a1/language-en>), p. 64.
- ¹³ 5x the amplification of a standard "like".
- ¹⁴ "Five points for anger, one for a 'like': How Facebook's formula fostered rage and misinformation", Washington Post, 26 October 2021 (URL: <https://www.washingtonpost.com/technology/2021/10/26/facebook-angry-emoji-algorithm/>).
- ¹⁵ "Digital Services Act: Application of the Risk Management Framework to Russian disinformation campaigns", European Commission, 30 August 2023 (URL: <https://op.europa.eu/en/publication-detail/-/publication/c1d645d0-42f5-11ee-a8b8-01aa75ed71a1/language-en>), p. 64.
- ¹⁶ CJEU judgement of 4 July 2023, *Bundeskartellamt v Meta*, C-252/21, ECLI:EU:C:2023:537, paragraph 102.
- ¹⁷ "Guidelines 05/2020 on consent under Regulation 2016/679", European Data Protection Board, 4 May 2020 (URL: https://edpb.europa.eu/sites/default/files/files/file1/edpb_guidelines_202005_consent_en.pdf), pp. 20-22.

Emailed to: VSPSregulation@cnam.ie

04 September 2023

Call For Inputs: Online Safety

Developing Ireland's First Binding Online Safety Code for Video-Sharing Platform Services

The Internet Commission (as part of the Trust Alliance Group) welcomes the opportunity to respond to the call for inputs regarding the development of Ireland's First Binding Online Safety Code for Video-Sharing Platform Services.

In our response we have provided:

Section 1: Introduction to the Trust Alliance Group and the Internet Commission.

Section 2: Answers to specific questions where we think we can contribute a helpful perspective.

Section 1 - Introduction to the Trust Alliance Group and the Internet Commission

Trust Alliance Group is a not-for-profit private limited company established in 2002 which runs a range of discrete national Alternative Dispute Resolution (ADR) schemes across different sectors, including the sole ADR scheme in the energy sector in England and Wales, the Ofgem-approved Energy Ombudsman and the Communications Ombudsman, approved by Ofcom.

The Internet Commission – a non-profit organisation which promotes ethical business practice to counter online harms whilst protecting privacy and freedom of expression and increase platform accountability – was acquired by the Trust Alliance Group in 2022.

The Internet Commission was conceived by Dr Ioanna Noula and Jonny Shipp in 2017 in the context of their research for the Department of Media and Communications at the London School of Economics where they were both visiting fellows. The drivers at the time for such research were various events from Cambridge Analytica and Facebook's interference in the US election to Molly Russell's suicide following her exposure to harmful content on Instagram.

Ioanna and Jonny gathered multiple stakeholders such as senior academics from LSE, UCL and Imperial College, government representatives (UK Government Digital Service, Future Cities Catapult) as well as business representatives like Siemens, Telefonica and Pearson Education. The aim was to discuss the impact of social media platforms' failure to self-regulate and the need for the development of checks and balances that would increase the accountability of digital service providers, safeguard citizens' rights and wellbeing online, and restore stakeholder trust in tech.

Subsequently, in 2018, the Internet Commission was founded, and started a round of digital responsibility assessments with prominent businesses which led to their first public accountability report in 2021. The Internet Commission offers:

- independent evaluation of online intermediaries (social media, news sites, dating service providers, gaming service providers, digital education providers etc.) regarding their practices of content moderation;
- knowledge exchange where companies can discuss challenges and solutions related to tackling online harms; and
- a bank of good practices and reporting on the state-of-the art regarding governance and procedures of moderation of user-generated content (UGC) online.

Our comments to this consultation come from our experience from evaluating global online service providers' platforms across different online services and consider the insight the Internet Commission has generated by taking a closer look at procedures, resources, governance and the organisations' culture driving UGC moderation. Our research has explored critical challenges faced by service providers such as:

- achieving maximum efficiency by balancing human and automated moderation;
- understanding the implications of outsourcing content moderation services;
- addressing tensions emerging from users' rights online (digital rights); and
- ensuring content moderators' wellbeing.

Specifically, we share evidence from our evaluation of a diverse cohort of online services including two dating service providers, a gaming service provider, a live-streaming gaming service provider, a news services organisation, and a children's social media service provider. We retain a focus on procedural accountability; that consumer outcomes, particularly vulnerable communities, are best served by ensuring that processes and procedures are evaluated, and we use this information to identify emerging trends and issues. Being proactive in this fast-moving space is key and our approach allows us to flex against market requirements.

Our independent evaluation takes a look "under the hood" at processes, culture and technology that shape content moderation and offer industry benchmarks UK wide and internationally.

Section 2 – Answers to Questions

Question 10: What requirements should the Code include about age verification and age assurance? What sort of content should be shown by default to users who are logged out or in private browsing mode and whose age cannot be verified or assured? What evidence is there about the effectiveness of age estimation techniques? What current practices do you regard as best practice? Where accounts are not age verified should default privacy settings be used, should content default to universal content and should contact by others be more limited?

As explained in the introduction, in 2020 the Internet Commission started a round of digital responsibility assessments with prominent businesses which led to our first public accountability report in 2021. We offered independent evaluations to social media platforms, news sites, dating and gaming service providers, and digital education providers etc. regarding their practices of content moderation. By doing so, we have identified various practices concerning age assurance across the two cohorts of businesses we assessed. All of the evidence concerning such practices has been collected and can be seen in Appendix 1 with both qualitative and quantitative data.

Practices identified during our analysis are placed within a maturity model, which begins at Stage I: Elementary and goes up to Stage V: Transforming. An age assurance practice that falls under Stage I is the implementation of an age gate that relies solely upon the self-declaration of age.

Self-declaration of age is entirely unreliable. Those users who are incentivised to lie about their age will do so, and will likely face little or few consequences for doing so - if they are ever found out. For many platforms, especially those where engagement drives profits through advertising, there has been very little commercial incentive to block users' access to the service based on age or to punish users if they evade what little blocks are in place.

We found out this underlying commercial reality with one of the platforms in our cohort which had such a business model and did not require any further age checks even when the content it hosts had been labelled 'mature' by its creator. On the other hand, it does operate 21+ age gates on channels featuring promotions of or sponsorships by alcohol brands. While these gates are still inadequate, being self-declared once again, they do demonstrate a responsiveness to more tightly regulated industries.

It is worth noting that the age gates on this platform were accompanied by temporary cookies which would be dropped to, for a short time, restrict a user's ability to create an account with another date of birth if they were blocked by that age gate. This sub-practice was again found to be an immature practice, particularly in comparison with other platforms.

For example, another member of the cohort also used self-declared age gates but supplemented this gate with additional tools to prevent users gaming the system and to build out a more holistic approach to age detection throughout the platform.

These (18+) age gates were buttressed by tools which - if a prospective user were to enter details that did not meet the requirements of the age gate - would lock those credentials until the user turned 18 and so were longer lasting than those used by the platform discussed above. The platform also deployed automated tools to detect underage users via photographs, biographies and private messages. Suspected underage users' accounts are suspended and can only be reinstated once their age has been verified as 18+ by a third-party service.

Crucially, this platform's business model was driven by paid subscriptions rather than advertising and engagement, such that it was not the case that all users were equally valuable to them and equally wanted on the platform. It was also built to facilitate real-life meetings between users and so there was much more of a commercial incentive for the platform to enforce its Terms of Service and ensure that the pool of users on the platform were of age.

In 2021, the accounts suspended for being underage as a proportion of all accounts suspended was 18% on the second platform. On the first platform, there is no single category including 'underage' as a reason for account suspension. It could only fall under 'other', which makes up around 54% of suspensions.

What has become clear through our work with companies and platforms catering to different demographics, and perhaps appealing to others, is that a one-size-fits-all approach is rarely appropriate.

On one platform designed for children, there wasn't a specific need for age assurance or age verification, despite there being an age limit of 13 and under in the Terms of Services, because it did not provide communication tools enabling users to privately communicate with one another.

While the idea of a child-only platform may raise concerns in that it would appear to be the ideal location for a predator or bad actor to operate within a walled garden, were they to circumvent whatever mechanism made it child-only, the absence of such a mechanism and means to privately communicate (alongside additional moderation tools) negated such a risk.

Other differences, including business model, as described above, and the outcome of use (e.g. in-person meetings), mean that varying [levels of assurance](#) should be required and applied to different platforms. This principle should equally be applied to non-users, or those whose ages cannot be verified.

Platforms should be equipped with the tools and rules to conduct an effective risk assessment of their platform and determine the requisite level of age assurance for each part of the user journey or segment of the platform.

Approaching the issue with a sensitivity to the differences between platforms will foster a dynamic ecosystem wherein platforms can comply without threatening to limit users' experiences and more closely approximate real-world approaches: for example, mirroring the kind of visual age estimation one might expect when buying a ticket to see a film in a cinema vs the more stringent checking of identity documentation when buying alcohol or, even more so, opening a bank account.

Question 15: How should we ask VSPS providers to address content moderation in the Code? Are there any current practices which you consider to be best practice? How should we address automated content detection and moderation in the Code?

In our experience, improvements to content moderation could be made by considering:

Moderator training and support

The moderation process should be respectful to users: when a post is removed, both the user that created the post and the “flagger” of the problem should be notified, with details of which content was removed, the rule broken and information about the appeals process. This could follow a well-developed process for broadcast television and radio, which includes a clear escalation path which dovetails with the established complaints process.

Quality Assurance

Appeals processes help get the balance right between safety and freedom of expression. Moderators and automated processes can remove too much or too little content. VSPS providers should hold regular quality assurance sessions where a sample of decisions can be checked, and feedback should be provided particularly on contentious issues should be part of a running dialogue in the organisation.

Quality assurance checking should ensure consistency across moderators at different periods of time. The number of appeals should also be tracked and evaluated by specialist quality assurance teams.

Integrated enforcement and appeals systems

Users need to be able to understand what activity causes a particular enforcement action to understand where they went wrong and be able to appeal if necessary. This also has impacts for moderation staff who must spend time checking across the two systems to validate the appeal. A disconnected approach may lead to questionable – or simply incorrect – moderation decisions. Moreover, educating the user through more transparency could minimise the impact of online activity that requires further sanctions.

Signposting mental health support

We are aware of a service provider who has partnered with a mental health service to signpost additional support to users who may benefit from such support. Users may text the name of the organisation to the mental health service provider to be connected with a counsellor immediately. It is also beneficial to consider mental health support and robust wellbeing programs for content moderators to ensure better outcomes.

Question 16: What requirements should the Code include about procedures for complaint-handling and resolution, including out-of-court redress and ADR? To what extent should these requirements align with similar requirements in the DSA? What current practices could be regarded as best practice? How frequently should VSPS providers be obliged to report to the Commission on their complaint handling systems and what should those reports contain? Should there be a maximum time-period for VSPS providers to handle user complaints and if so, what should that period be?

We work with a range of service providers - including a dating service provider, gaming service provider, news services, and children's social media service providers. We have seen service providers implement different ways in which they enhance the transparency, accessibility and awareness of reporting and complaint mechanisms. These include:

- Ensuring there is a formal right of appeal process and that it is clear to users and available to non-users (especially important in relation to the parents of users).
- Sharing details of which content has been identified as inappropriate or harmful and information on the appeals process. This approach aims to treat users as trustworthy contributors, with a focus first on users' intentions when reaching a judgement about the suitability of their posts.
- Apology mechanisms that are followed for users which have been found via the appeals process to have been wrongfully banned. This can encourage a shared sense of accountability.
- Progress updates on appeals and, in the case of one organisation, a forthcoming dashboard for appeals which will allow for integration of the enforcement and appeals systems. While it may appear that this would be necessary for proper functioning and naturally happen, the staggered development of systems can lead to nonconformity between them. It should be at the very least recommended, then, that enforcement and appeals should be linked at the back end to facilitate more effective decision-making processes for moderators and greater clarity for users.

The reporting routes for children, as opposed to adults, are not currently clear in the sector but some providers are looking at simplifying their appeals process to make it more accessible to vulnerable groups. We believe this is an important step and are keeping this area under review.

Call For Inputs: Online Safety: Developing Ireland's First Binding Online Safety Code for Video-Sharing Platform Services

Submission from Brian O'Neill, PhD, Researcher, Emeritus Professor, TU Dublin.

I would like to thank Coimisiún na Meán for the opportunity to respond to its call for inputs on *Developing Ireland's First Binding Online Safety Code for Video-Sharing Platform Services*. Bringing forward the first online safety code will be a highly significant milestone in policy development on this topic with important implications not just for online safety in Ireland but for digital policy more generally, nationally and across the EU.

As an academic researcher in the area of children and the digital environment, the following observations draw on my experience of participating in initiatives such as EU Kids Online, a multinational research network that seeks to enhance knowledge of European children's online opportunities, risks and safety, and other similar research networks. I have also contributed to the Council of Europe's Digital Citizenship Education initiative, which embeds the values of human rights, democracy and the rule of law in digital literacy education. My current work involves contributing to policy mapping in relation to the European Commission's Better Internet for Kids (BIK+) strategy, which, in part, informs some of the observations below. I should also note that I serve as Deputy Chair of the National Advisory Council for Online Safety (NACOS) (and contributed to its research study *Report of a National Survey of Children, their Parents and Adults regarding Online Safety 2021*). This submission is made in a personal capacity.

Question 1: What do you think our main priorities and objectives should be in the first binding Online Safety Code for VSPS? What are the main online harms you would like to see it address and why?

The first binding Online Safety Code for VSPS is an important statement of intent regarding online safety and will likely set a benchmark for future policy instruments. As such, the Code should aim to be as comprehensive as possible, future-proofed and based on a solid foundation of international standards vis à vis rights and responsibilities in the digital environment.

The Code specifically focuses on VSPS and serves a particular function under Article 28b AVMSD. VSPS have a distinctive and evolving role in the social media ecosystem. The principles for protecting minors and online safety to be articulated in the Code will likely act as a template for future regulatory statements on this topic.

The code should take account of the deliberations on the *EU Code of Conduct on age-appropriate design*,¹ which, as proposed within the BIK+ strategy,² will be in line with AVMSD and GDPR and will build on the rules of the DSA. Coimisiún na Meán is one of the first designated Digital Services Coordinators (DSC) under the Digital Services Act. Given that Irish-hosted digital services come within its remit, the cross-border implications are significant.

At a minimum, the online harms set out in national legislation and the AVMSD should be mandatory for all providers. However, rather than specifying particular categories of online harm, the emphasis should be outcomes-based so that the effect is systemic and based on appropriate risk assessment.

Ensuring a robust and principles-based foundation for the Code should be a priority to underpin its systemic and sector-wide reach. The UN Committee on the Rights of the Child *General comment No.*

¹ <https://digital-strategy.ec.europa.eu/en/news/crafting-code-conduct-age-appropriate-design-kicks-today>

² <https://digital-strategy.ec.europa.eu/en/policies/strategy-better-internet-kids>

25 (2021) on children's rights in relation to the digital environment³ is a valuable statement in this regard, as are the Council of Europe's *Guidelines to respect, protect and fulfil the rights of the child in the digital environment - Recommendation CM/Rec(2018)7 of the Committee of Ministers (2018)*⁴ and the OECD's *Guidelines for Digital Service Providers (2021)*.⁵ An important example of a rights-based code at a national level is the Dutch *Code voor Kinderrechten (2021)* (Code for Children's Rights).⁶

Question 2: What types of online harms do you think should attract the most stringent risk mitigation measures by VSPS? How could we evaluate the impact of different kinds of harm, e.g. severity, speed at which harm may be caused? Is there a way of classifying harmful content that you consider it would be useful for us to use?

The most stringent risk mitigation measures should be those proscribed in law and which are the subject of international conventions and EU regulations covering such areas as child sexual abuse content, pro-terror content and illegal hate speech. Combatting illegal content and online activity remains the top priority for all stakeholders, for which the Code acts as a further important instrument.

While illegal content is the priority, dealing with content that is harmful but not illegal is one of the challenging issues that the Code needs to address. Classifying content that may be harmful but not illegal relies on judgements that may be context-specific and which risk curtailing rights to freedom of expression. For this reason, AVMSD has consistently maintained a graduated approach⁷ towards regulating content on services, with common rules in line with EU laws and stricter regulations scaled according to the degree of severity and likely impact on minors.

In summary, as a priority, illegal content (offence-specific categories of online content), as defined in national and EU law, should attract the most stringent risk mitigation measures. The Code should then give effect to measures addressing other harmful content that meet the high bar of the risk test in OSMR, i.e., where there is a risk to a person's life or poses a significant risk of harm to a person's mental/physical health, which is reasonably foreseeable.

OSMR further specifies harmful content as consisting of one of the following: content of one person bullying or humiliating another; content which promotes or encourages eating disorders; and content which promotes or encourages self-harm/ suicide or makes available information on methods of self-harm/suicide. These should be subject to risk mitigation measures in accordance with the likelihood of access and potential impact on vulnerable subjects.

By way of illustrating the challenge in delineating and codifying harmful content, Australia's eSafety Commission refers to the national classification scheme for harmful online content to support its development of industry codes under the Online Safety Act (2021).⁸ The classification of so-called Class 1 and Class 2 material, as defined under a scheme for the classification of films, publications and computer games, is used to define obligations for service providers. Class 1 material refers to extreme content that would be refused classification under the national scheme, the production and possession of which is legally proscribed. Class 2 materials are those that are likely to be restricted under Australia's national classification scheme and for which there is evidence that it may cause harm to vulnerable groups. A difficulty with this approach is that it incorporates illegal, harmful and offensive content, thereby creating ongoing challenges in delineating where boundaries occur. While the classification scheme is currently under review, the eSafety Commissioner has acknowledged its suitability to the online environment is limited, particularly as it was initially developed for commercially

³ <https://www.ohchr.org/EN/HRBodies/CRC/Pages/GCChildrensRightsRelationDigitalEnvironment.aspx>

⁴ <https://edoc.coe.int/en/children-and-the-internet/7921-guidelines-to-respect-protect-and-fulfil-the-rights-of-the-child-in-the-digital-environment-recommendation-cmrec20187-of-the-committee-of-ministers.html>

⁵ <https://www.oecd.org/mcm/OECD%20Guidelines%20for%20Digital%20Service%20Providers.pdf>

⁶ <https://codevoorkinderrechten.nl/>

⁷ <https://digital-strategy.ec.europa.eu/en/policies/general-principles-avmsd>

⁸ eSafety Commissioner (2021). Development of industry codes: position paper. Available at: <https://www.esafety.gov.au/sites/default/files/2021-09/eSafety%20Industry%20Codes%20Position%20Paper.pdf>

produced rather than user-generated content and requires considerable input on the part of the regulator to assess its implications for incorporation into industry codes.

Question 3: Do you have reports, academic studies or other relevant independent research that would support your views? If you do, please share them with us with links to relevant reports, studies or research.

EU Kids Online has published a number of research studies related to the classification of online risks for children. Its 3Cs classification of risks – Content, Contact and Conduct risks (later expanded to include a 4th C of Commercial risks) was first published in 2011 and has been widely influential. Importantly, EU Kids Online has always underlined that ‘risk’ is the *probability* but not the *inevitability* of ‘harm’. Hence, the importance of risk mitigation and resilience measures.⁹

Relevant publications and studies include:

Stoilova, M., Rahali, M., & Livingstone, S. (2023). *Classifying and responding to online risk to children: Good practice guide*. Insafe helplines and the London School of Economics and Political Science (LSE). <https://www.lse.ac.uk/business/consulting/assets/documents/Classifying-and-responding-to-online-risk-to-children-Good-practice-guide.pdf>

Livingstone, S., & Stoilova, M. (2021). The 4Cs: Classifying Online Risk to Children. *CO:RE Short Report Series on Key Topics*. <https://doi.org/10.21241/SSOAR.71817>

O’Neill, B. (2023). Research for CULT Committee – The influence of social media on the development of children and young people. European Parliament, Policy Department for Structural and Cohesion Policies, Brussels. <https://bit.ly/3XkgYd8>

Question 4: What approach do you think we should take to the level of detail in the Code? What role could non-binding guidance play in supplementing the Code?

As described in the call for inputs, a mixed approach with high-level obligations as the most prominent feature of the Code, supported by more concrete guidance as required, would appear to be the most appropriate.

In order to meet the objective of achieving long-term systemic change with online safety and safety-by-design moving centre stage in the development of digital services, some flexibility is needed in the Code. A very detailed or prescriptive code would lack this flexibility, act as a disincentive to innovating for online safety and place an unnecessary burden of responsibility on Coimisiún na Meán to assess risks in the digital environment.

At the same time, Coimisiún na Meán can fulfil an important leadership function in guiding the sector towards improved online safety standards by developing appropriate independent guidance. The Data Protection Commission’s *Children Front and Centre: Fundamentals for a Child-Oriented Approach to Data Processing* (2021) is a good example of how this can work in practice.

Question 5: What do you think would be the most effective structure for the Code? What are the most important factors we should consider when we decide how to structure the Code?

Following the suggestion of a mixed approach with a focus on high-level obligations, the most appropriate structure would appear to be a thematic one and organised around relevant sections

⁹ See S. Livingstone (2021). “More online risks to children, but not necessarily more harm: EU Kids Online 2020 survey”. Available at: <https://blogs.lse.ac.uk/medialse/2020/02/11/more-online-risks-to-children-but-not-necessarily-more-harm-eu-kids-online-2020-survey/>

dealing with Content Policies / T&Cs, Risk Assessments, Content Moderation and Complaints, Online Safety Features, Service Design Measures, Compliance Measures etc.

This has the advantage of maximising transferability to other contexts and maintaining consistency across the sector.

A useful model to consult is the template or 'preferred codes model' developed by Australia's eSafety Commissioner in its position paper on industry codes.¹⁰

Question 6: How should we design the Code to minimise the potential for conflict and maximise the potential for synergies in how platforms comply with it and the DSA?

There must be continuity across the respective regimes and compliance requirements for all applicable laws and regulations to reinforce a consistent message regarding online safety standards and encourage the highest levels of compliance. Mirroring provisions of the DSA make sense in this context, particularly those relating to VLOPs, which are likely to attract significant attention. As noted under Q.1, liaising closely with the *EU Code of Conduct on age-appropriate design* (in development) would be beneficial as it is intended to operate within the DSA's rules.

Question 7: To what extent, if at all, should the Code require VSPS providers to take measures to address content connected to video content?

Measures specified under the Code should address all related content, not just the depictions in a video stream. VSPS place a lot of emphasis on the integrated user experience on their platforms, all of which contribute to the sometimes highly complex and multi-dimensional nature of the communication. This complexity already forms part of the content moderation process for many platforms, including within their T&Cs community rules governing all aspects of content shared on the site. Accordingly, obligations set out under the Code should reflect this reality and require a holistic approach by providers in providing a safe online environment.

It would also be beneficial if transparency reports of moderation decisions produced by providers include details of where infringements occur. This important data can contribute to a better understanding of user behaviours and spotlight design weaknesses and areas requiring greater attention for risk assessment.

Question 8: How should we ask VSPS providers to introduce a feature that allows users to declare when videos contain advertising or other types of commercial communications? Should the Code include specific requirements about the form in which the declaration should take? What current examples are there that you regard as best practice?

The importance of clear communication regarding content containing advertising and other types of commercial communications is vitally important. The lack of legibility and the blurred boundaries surrounding embedded commercial content is something that researchers have highlighted as a persistent challenge in children's consumption of social media content. The EU Kids Online network updated the classification of online risks in 2021 to include a fourth 'C' of "contract risks" to reflect the specific issues posed by commercialisation and datafication and to reflect the many profound changes that have taken place in the digital environment since the typology was first created. The OECD also added "consumer risks" to its typology of online risks in its updated Recommendation on Children in the Digital Environment¹¹ to convey the wide range of contexts in which children are exposed to online commercialised messaging and for which they may be ill-prepared.

¹⁰ Chapter 5 'Preferred codes model', *Development of industry codes under the Online Safety Act Position Paper*. Available at: <https://www.esafety.gov.au/sites/default/files/2021-09/eSafety%20Industry%20Codes%20Position%20Paper.pdf>

¹¹ <https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0389>

Evolving marketing strategies, targeted advertising, and personalised profiling are frequently opaque and pose ever-increasing challenges for users in identifying commercial content. These have been deemed as unfair practices which exploit children's incredulity and lack of experience while significantly impacting their rights and well-being (Hof et al., 2020).¹² Content scaled for smartphones and mobile devices is a further challenge because cues signalling commercial content are even harder to see.¹³ Advertising literacy is particularly important for younger children for whom video-sharing platforms are a key part of their media consumption. A study of preschool children aged 4 to 5 years in Flanders showed that even with appropriate tagging, most children displayed no critical advertising literacy, treating the advertisement the same way as the entertainment content (Vanwesenbeeck et al., 2020).¹⁴

Concerning solutions, as with other aspects of online safety, a combination of approaches, including education, are needed. Empirical studies have shown that the prominence of tags matters, as argued by the ICCP. This is particularly the case with influencer-based marketing¹⁵ widely used on VSPS, which have been the subject of several European regulatory interventions. In 2022, the Spanish media regulator, CNMC, introduced new rules for vloggers and online influencers requiring greater transparency with plans to establish a State Registry of Audiovisual Communication Service Providers.¹⁶

Question 9: How should we ask VSPS providers to introduce and design a flagging mechanism in the Code? How can we ensure that VSPS providers introduce the mechanism in a user-friendly and transparent way? How should we ask VSP Providers to report the decisions they've made on content after it has been flagged? To what extent should we align the Code with similar provisions on flagging in the DSA?

There is an emerging international consensus on standards that should apply in the design of online safety features that conform to principles of Safety by Design (SbD), as evidenced, for example, by the UK government's guidance *Principles of safer online platform design*,¹⁷ and the *Age appropriate design: a code of practice for online services*.¹⁸ Industry codes and guidelines have also addressed design issues, e.g., the Digital Trust & Safety Partnership *Best Practices Framework*.¹⁹

More work needs to be done on standardisation in this area. The IEEE Standard for an *Age Appropriate Digital Services Framework Based on the 5Rights Principles for Children* offers an overview of how design standards might apply.²⁰ A valuable resource regarding design issues for online safety features such as reporting mechanisms is the series of materials on SbD published by Australia's eSafety Commissioner.²¹ This includes tools aimed at companies for assessing how systems, processes and practices support user safety based on principles and good practice in SbD.

¹² Hof, S. van der, Lievens, E., Milkaite, I., Verdoodt, V., Hannema, T., & Liefwaard, T. (2020). The Child's Right to Protection against Economic Exploitation in the Digital World. *The International Journal of Children's Rights*, 28(4), 833–859. <https://doi.org/10.1163/15718182-28040003>

¹³ Feijoo, B., & Sádaba, C. (2022). When Ads Become Invisible: Minors' Advertising Literacy While Using Mobile Phones. *Media and Communication*, 10(1), Article 1. <https://doi.org/10.17645/mac.v10i1.4720>

¹⁴ Vanwesenbeeck, I., Hudders, L., & Ponnet, K. (2020). Understanding the YouTube Generation: How Preschoolers Process Television and YouTube Advertising. *Cyberpsychology, Behavior, and Social Networking*, 23(6), 426–432. <https://doi.org/10.1089/cyber.2019.0488>

¹⁵ Van Reijmersdal, E. A., Rozendaal, E., Hudders, L., Vanwesenbeeck, I., Cauberghe, V., & Van Berlo, Z. M. C. (2020). Effects of Disclosing Influencer Marketing in Videos: An Eye Tracking Study among Children in Early Adolescence. *Journal of Interactive Marketing*, 49(1), 94–106. <https://doi.org/10.1016/j.intmar.2019.09.001>

¹⁶ <https://www.boe.es/boe/dias/2022/07/08/pdfs/BOE-A-2022-11311.pdf>

¹⁷ <https://www.gov.uk/guidance/principles-of-safer-online-platform-design>

¹⁸ <https://ico.org.uk/media/for-organisations/guide-to-data-protection/key-data-protection-themes/age-appropriate-design-a-code-of-practice-for-online-services-2-1.pdf>

¹⁹ <https://dtspartnership.org/>

²⁰ <https://ieeexplore.ieee.org/document/9627644/>

²¹ <https://www.esafety.gov.au/industry/safety-by-design>

Question 10: What requirements should the Code include about age verification and age assurance? What sort of content should be shown by default to users who are logged out or in private browsing mode and whose age cannot be verified or assured? What evidence is there about the effectiveness of age estimation techniques? What current practices do you regard as best practice? Where accounts are not age verified should default privacy settings be used, should content default to universal content and should contact by others be more limited?

Art28b3 AVMSD requires that VSPS should establish and operate age verification systems for users of their platforms with respect to content that may impair the physical, mental or moral development of minors. The details of such systems are not specified, and it is the case that age verification systems vary in terms of sophistication, effectiveness and compatibility with data protection requirements. The technical solutions continue to evolve at a rapid pace but still pose challenges as regards suitability. There is, however, an emerging effort to build sector-wide interoperability, as illustrated, for example, by the euConsent project.²²

Notwithstanding these constraints, it is for industry providers to demonstrate that adequate safeguards are in place to ensure that content "might seriously impair" the development of minors is not accessible, i.e., that age assurance goes beyond the self-declaration methods that have primarily applied to date. As demonstrated by the Italian DPC's action against TikTok, where such obligations are made explicit, system improvements follow.²³

Concerning default settings, the Irish DPC's requirement that a floor of protection applies where accounts are not age-verified is an important principle to follow (Principle 1 – Fundamentals).²⁴

Question 11: What requirements should the Code have in relation to content rating? What do you consider to be current best practice? What experiences have you had using content rating systems on platforms and do you think they have been effective? What steps could we ask VSPS to take to ensure content is rated accurately by users?

A requirement in the Code that VSPS establish and operate easy-to-use systems that allow users to age-rate the videos they upload would be a valuable boost to this online safety feature. Experience to date with labelling systems for online content is mixed. As noted in the call for inputs, there are similarities across the many existing rating systems, but there are also many variations due to the many social and cultural differences involved. The comparison with film classification schemes is also not an exact one as noted by the Australian eSafety Commissioner. In online and user-generated content, schemes such as PEGI and PEGI online²⁵ may be closer to the VSPS context, particularly regarding the processes followed to rate the content.

Wider use of content classification was one of the priority themes addressed decade ago by the CEO Coalition self-regulatory initiative overseen by the European Commission.²⁶ One of the outcomes of this process was the You Rate It system, coordinated by the highly experienced classification bodies NICAM and BBFC.²⁷ The system has struggled to gain traction, however, partly because platforms have little incentive or obligation to use it. It is, however, a tailor-made solution and merits consideration. The IFCO was an early partner in the consortium.

The MIRACLE project is another example of an approach towards European standardisation and interoperability of age classification systems.²⁸

²² <https://euconsent.eu/>

²³ <https://www.garanteprivacy.it/web/guest/home/docweb/-/docweb-display/docweb/9524224>

²⁴ <https://www.dataprotection.ie/en/dpc-guidance/fundamentals-child-oriented-approach-data-processing>

²⁵ <https://pegi.info/page/pegi-online>

²⁶ <https://digital-strategy.ec.europa.eu/en/library/ceo-coalition-2014-progress-reports-actions-make-internet-better-place-kids>

²⁷ <https://www.yourateit.eu/>

²⁸ <https://leibniz-hbi.de/en/projects/miracle-interoperable-age-classifications>

Question 12: What requirements should the Code have in relation to parental control features? How can we ensure that VSPS providers introduce the mechanism in a user-friendly and transparent way? Can you point to any existing example of best practice in this area? Should parental controls be ‘turned-on’ by default for accounts of minors or where age is not verified?

Wider availability and use of parental controls have also long been advocated as an essential online safety measure (endorsed by the CEO Coalition and now-defunct ICT Coalition). Parental controls are also a feature of mobile operators’ codes of conduct, as featured in the *European Framework for Safer Mobile Use by Younger Teenagers and Children*.²⁹ Technology and usage trends, however, have arguably made reliance on parental controls obsolete. As a result, there is a trend towards systems that foster communication and dialogue between parents and children rather than a blanket access control mechanism.

Researchers and child rights advocates have also called attention to the potential for parental control systems to conflict with children’s rights to autonomy and access to information.³⁰ While a case may be made for a greater need for parental controls to operate with younger users, this is also dependent on the nature of the service and the child’s age. A survey undertaken for the euConsent project elaborates on the outcomes for children and families through parental controls as a child protection measure and contains recommendations for design practice.³¹ This is also relevant to the design of age assurance systems.

Question 13: What requirements should the Code contain to ensure that VSPS provide effective media literacy measures and tools?

Media literacy – the ability to understand and critically evaluate broadcast, online and other media content and services – is, as Media Literacy Ireland argues,³² a pre-requisite for citizenship in the digital age. No longer an add-on or a complementary educational measure to support users’ knowledge and skills, it is necessary in today’s complex, digitally-saturated information environments.

The necessity for media literacy arises from the very nature of risks within the digital environment and over which digital service providers have significant responsibility. Many providers have supported media literacy initiatives and organisations as part of their corporate social responsibility. Arguably, there is a need to do more and to build in – as envisaged by the call for inputs – media literacy tools and measures within the core functions of the platform and as part of the process of serving content to users.

Hence, there is an opportunity in the Code to require more concrete action by services that reinforces users’ media literacy. This might be carried through appropriate notifications, flags, posts, and in-feed prompts flagging, for example, unverified content, possible disinformation, links to in-platform tools and resources, and links to external fact-checking services and media literacy organisations. Warnings and labels, when designed well, have been found to help users identify and avoid disinformation.³³

²⁹ <https://www.gsma.com/gsmaeurope/safer-mobile-use/european-framework/>

³⁰ Zaman, B., & Nouwen, M. (2016). *Parental controls: Advice for parents, researchers and industry*. EU Kids Online. <http://eprints.lse.ac.uk/id/eprint/65388>

³¹ Smirnova, S., Livingstone, S., & Stoilova, M. (2021). *Understanding of user needs and problems: A rapid evidence review of age assurance and parental controls* (p. 60). euCONSENT. <https://euconsent.eu/download/understanding-of-user-needs-and-problems-a-rapid-evidence-review-of-age-assurance-and-parental-controls/>

³² <https://www.medialiteracyireland.ie/>

³³ Kaiser, B., Wei, J., Lucherini, E., Lee, K., Matias, J. N., & Mayer, J. (2021). *Adapting Security Warnings to Counter Online Disinformation*. 1163–1180. <https://www.usenix.org/conference/usenixsecurity21/presentation/kaiser>

Question 14: How should we ask VSPS providers to address online harms in their terms and conditions in the Code, including the harms addressed under Article 28b? How should key aspects of terms and conditions be brought to users' attention? What examples are there of best practice in relation to terms and conditions including content moderation policies and guidelines?

Terms and conditions that are comprehensive but also transparent, easy to read and accessible to users are an essential feature of platform online safety. The DSA requires wide-ranging transparency measures for online platforms, including better information on terms and conditions and transparency on the algorithms used for recommending content or products to users.³⁴ The BIK+ strategy also states that: “age-appropriate, easily understandable and accessible information, such as terms and conditions, instructions and warnings, and simple mechanisms to report harm should accompany all products and services likely to be used by children”.³⁵

Methods to ensure easy-to-read terms include relevant community rules or guidelines tailored to the needs of different age groups supported by help articles and resources in a clearly identifiable Safety Centre or equivalent. This is also an essential opportunity for platforms to elaborate on their policies and responses to risks and harms that may be particularly relevant to their service, including appropriate content moderation policies and guidelines.

Drawing from the field of consumer marketing, a UK government-commissioned best practice guide points to practical steps companies can take to improve users' understanding of contractual terms and platform policies and guidelines. These include displaying key terms as frequently asked questions, using icons to illustrate critical terms, providing information in short chunks at the right time, and telling users how long it will take to read a policy.³⁶

Question 15: How should we ask VSPS providers to address content moderation in the Code? Are there any current practices which you consider to be best practice? How should we address automated content detection and moderation in the Code?

Content moderation systems and processes are the bedrock of platforms' risk management approach, and as such, clear and comprehensive information about the quality and capacity of platforms' systems is vital. eSafety's SbD Principles provide a valuable overview of expectations for robust and effective implementation (as further elaborated, for instance in series of implementation reports on its Basic Online Safety Expectations – BOSE – process).³⁷ The latter includes examples of reasonable steps a provider may be expected to take in dealing with a range of online risks.

The Santa Clara Principles on Transparency and Accountability in Content Moderation are also a valuable resource for benchmarks and standards in this area.³⁸ Drafted by a range of academics and civil society organisations and endorsed by major companies in the technology sector, the Principles have been effective in bringing about greater consistency and driving change in the quality of information regarding moderation processes. Now in its second edition, the Santa Clara Principles 2.0 set a standard for transparency and accountability, for instance, requiring companies to “publish clear and precise rules and policies relating to when action will be taken concerning users' content or accounts, in an easily accessible and central location” (Foundational Principle #2) including “a comprehensive understanding of companies' processes and systems requires transparency around the use of automated decision-making tools” (Operational Principle #1).

³⁴ https://ec.europa.eu/commission/presscorner/detail/en/QANDA_20_2348

³⁵ *A Digital Decade for children and youth: the new European strategy for a better internet for kids (BIK+)*, p.10.

³⁶ Behavioural Insights Team. (2019). *Best practice guide. Improving consumer understanding of contractual terms and privacy policies: Evidence-based actions for businesses*. https://www.bi.team/wp-content/uploads/2019/07/BIT_WEBCOMMERCE_GUIDE_DIGITAL.pdf

³⁷ <https://www.esafety.gov.au/industry/basic-online-safety-expectations/responses-to-transparency-notice>

³⁸ *Santa Clara Principles on Transparency and Accountability in Content Moderation*. Retrieved 20 August 2023, from <https://santaclaraprinciples.org/>

Question 16: What requirements should the Code include about procedures for complaint-handling and resolution, including out-of-court redress or alternative dispute resolution processes? To what extent should these requirements align with similar requirements in the DSA? What current practices could be regarded as best practice? How frequently should VSPS providers be obliged to report to the Commission on their complaint handling systems and what should those reports contain? Should there be a maximum time-period for VSPS providers to handle user complaints and if so, what should that period be?

Complaints handling in this context refers to appeals and complaints processes related to content moderation decisions. As a core requirement of Art 28b AVMSD, i.e., the operation of “transparent, easy-to-use and effective procedures for the handling and resolution of users' complaints to the video-sharing platform provider in relation to the implementation of the measures” (i).

This is separate to the reporting function on platforms which itself is a core element of the content moderation process, e.g., through user reporting or flagging of content that may be in breach of a platform's terms and conditions. In the case of the latter, Principle 1.3 of the Safety by Design Overview (Australian eSafety Commissioner) states that providers should “Put in place infrastructure that supports internal and external triaging, clear escalation paths and reporting on all user safety concerns, alongside readily accessible mechanisms for users to flag and report concerns and violations at the point that they occur”.

Effective complaints handling is essential to preserving trust and transparency of the content moderation processes operated by the platform and an essential element of the fundamental rights to freedom of expression involved. As with the relevant terms of service, clarity in respect of the rationale for decisions on appeals and complaints is vital. Notably, successive OECD benchmarking reports regarding transparency reporting of the global top-50 online content sharing services found this to be highly inconsistent and lacking in transparency.³⁹

Relatedly, the Report of the Expert Group on an Individual Complaints Mechanism⁴⁰ deals with complaints handling processes, specifically in the context of how the public might complain to an Online Safety Commissioner about individual items of content that they suspect may fall within a category of harmful online content. This is potentially relevant to the consideration of alternative dispute resolution processes.

Question 17: What approach do you think the Code should take to ensuring that the safety measures we ask VSPS providers to take are accessible to people with disabilities?

A hallmark of good practice in safety by design is that safety is a primary consideration from the start rather than retrofitted or an afterthought, is user-centred and considers access by all groups, including those with disabilities. Principles of universal design come into play here so that the design and composition of an environment – including an online environment – is “can be accessed, understood and used to the greatest extent possible by all people regardless of their age, size, ability or disability” (Disability Act 2005).⁴¹

Question 18: What approach do you think the Code should take to risk assessments and safety by design? Are there any examples you can point us towards which you consider to be best practice?

³⁹ OECD (2022), “Transparency reporting on terrorist and violent extremist content online 2022”, *OECD Digital Economy Papers*, No. 334, OECD Publishing, Paris, <https://doi.org/10.1787/a1621fc3-en>.

OECD (2023), “Transparency reporting on child sexual exploitation and abuse online”, *OECD Digital Economy Papers*, No. 357, OECD Publishing, Paris, <https://doi.org/10.1787/554ad91f-en>.

⁴⁰ <https://www.gov.ie/en/publication/a7d97-report-of-the-expert-group-on-an-individual-complaints-mechanism-may-2022/>

⁴¹ See <https://universaldesign.ie/what-is-universal-design/>

See also Question 9 above, in particular the Australian eSafety Commissioner's *Safety by Design Overview*.⁴² This includes descriptions and good practices of documented risk management and impact assessments to assess and remediate any potential safety harms that could be enabled or facilitated by digital products or services (Principle 1.6). Risk assessment approaches are specific to individual services but also have in common a commitment to assessing risk for each feature and architectural component before it is brought to market.

Central to children's online safety in this context is the notion of a child rights impact assessment⁴³ which, over and above safety risk assessments, examines the potential impact on the full spectrum of children's rights. The BIK+ strategy, for example, contains three pillars of safe online use or *protection, digital empowerment* and *active participation* – each of which are needed to ensure children get the most out of the digital environment. UNICEF has developed its *MO-CRIA: Child Rights Impact Self-Assessment Tool for Mobile Operators*⁴⁴ in the context of digital products and services for mobile devices. The Digital Futures Commission (5Rights Foundation) has developed a dedicated resource called *Child Rights by Design* containing guidance for innovators when designing digital products and services.⁴⁵ A separate *Playful by Design* is a toolkit to support designers improve children's opportunities for free play in a digital world, and to tackle the challenges in developing digital products and services that respect children's rights.⁴⁶

Question 19: How do you think that cooperation with other regulators and bodies can help us to implement the Code for VSPS?

The importance of cooperation between regulators dealing with distinct aspects of digital services is well-established. The Digital Regulators Group established previously by the Broadcasting Authority of Ireland (BAI), Competition and Consumer Protection Commission (CCPC), Commission for Communications Regulation (ComReg) and the Data Protection Commission (DPC) has demonstrated the value of sharing knowledge and expertise and ensuring a consistent response to a complex and evolving digital environment.

Similar approaches exist in the UK (the Digital Regulation Cooperation Forum)⁴⁷ and in Australia (the Digital Platform Regulators' Forum)⁴⁸ for similar reasons of supporting a streamlined and cohesive approach to the regulation of digital platforms.

Within the EU, this takes on particular significance given the need for cooperation to give effect to the country of origin principle. The European Board for Digital Services will have specific functions to support the consistent application of the DSA just as the European Regulators Group for Audiovisual Media Services (ERGA) has played a key role in the development of codes of practice on disinformation.

The creation of the Global Online Safety Regulators Network⁴⁹ has similarly been valuable for knowledge exchange, particularly given the early stage of development of regulatory practice in this area.

⁴² <https://www.esafety.gov.au/sites/default/files/2019-10/SBD%20-%20Overview%20May19.pdf>

⁴³ https://sites.unicef.org/csr/css/Children_s_Rights_in_Impact_Assessments_Web_161213.pdf

⁴⁴ <https://www.unicef.org/media/97371/file/MO-CRIA:%20Child%20Rights%20Impact%20Self-Assessment%20Tool%20for%20Mobile%20Operators.pdf>

⁴⁵ https://digitalfuturescommission.org.uk/wp-content/uploads/2023/03/CRbD_report-FINAL-Online.pdf

⁴⁶ <https://digitalfuturescommission.org.uk/playful-by-design-toolkit/>

⁴⁷ <https://www.gov.uk/government/collections/the-digital-regulation-cooperation-forum>

⁴⁸ <https://www.esafety.gov.au/about-us/consultation-cooperation/digital-platform-regulators-forum>

⁴⁹ <https://www.esafety.gov.au/about-us/who-we-are/international-engagement/working-international-forums-and-networks>

Question 20: What approach do you think we should take in the Code to address feeds which cause harm because of the aggregate impact of the content they provide access to? Are there current practices which you consider to be best practice in this regard?

Online harm is complex and rarely attributable to one factor or one piece of content. This is reflected in the EU Kids Online conceptual model, which looks at various factors such as the individual's skills and psychological resources, quality of mediation supports and the conditions of access that may determine how and when harm occurs and to which degree of severity.⁵⁰ Similarly, the cross-platform nature of much online harm and offence is such that diverse pieces of content, which individually may not be in breach of a platform's rules or deemed harmful in their own right, have untold pernicious effects.

This reminds us that there is no magic bullet solution to complex issues of online harm and that platforms need to adopt a broad overarching duty of care to their users while there is also a shared responsibility for stakeholders to contribute to online safety. The commitment within codes of practice towards media literacy should also be taken to include support for building resilience and user empowerment, particularly in recognising signals of potential problems and helping those who may be vulnerable to seek appropriate support.

Question 21: Do you have any views on how requirements for commercial content arranged by a VSPS provider itself should be reflected in the Code?

Section 46n of the OSMR Act provides for setting standards that govern commercial communications to protect the interests of the audience, and where they relate to children, protect their interests in particular with regard to their general public health. The BAI Children's Commercial Communications Code has been an effective instrument in this regard and there is a case for a new commercial code to address the changed circumstances for marketing and commercial content. While the DSA forbids profiling children for targeted advertising, social media and VSPS remain thoroughly commercial environments using diverse overt and less overt communication methods that may exploit children's vulnerability. A roundtable hosted as part of the consultations on the BIK+ strategy deliberated on this topic and called for greater inter-agency cooperation and future-proofing of standards given the fast pace of change in technology platforms and commercial practices.⁵¹

Question 22: What compliance monitoring and reporting arrangements should we include in the Code?

With regard to transparency reporting requirements, this in tandem with equivalent requirements under DSA, is an area where the Code can set specific expectations.

Transparency and accountability reporting have been the subject of international policy debate with a number of proposals on how to improve practice. In 2022, the OECD launched its Voluntary Transparency Reporting Framework (VTRF), a web portal for submitting and accessing standardised transparency reports from online content-sharing services about their policies and actions on terrorist and violent extremist content (TVEC) online.⁵² Using a standardised questionnaire that covers 12 main topics, the framework is designed to be answerable by services of all sizes and intended to produce a baseline level of transparency. Benchmarking reports for the world's top 50 online content

⁵⁰ Livingstone, S., Mascheroni, G., & Staksrud, E. (2015). *Developing a framework for researching children's online risks and opportunities in Europe* [Monograph]. EU Kids Online, The London School of Economics and Political Science. <http://eprints.lse.ac.uk/64470/>

⁵¹ EUN. (2022). *Roundtable on child and youth consumer protection in digital markets: Roundtable Report*. European Schoolnet. <https://www.betterinternetforkids.eu/documents/167024/6966559/Roundtable+on+child+and+youth+consumer+protection+in+digital+markets+-+Report+-+FINAL+-+December+2022.pdf/cfa690fa-30d5-4480-0b12-269184b3a047?t=1670755374414>

⁵² <https://www.oecd.org/digital/vtrf/>

sharing services have been published in respect of TVEC⁵³ and shortly in relation to online child sexual abuse and exploitation (forthcoming).⁵⁴

Also in 2022, the industry alliance, the Tech Coalition, launched its *Trust: Voluntary Framework for Industry Framework*. This sets a set of minimum requirements in transparency reporting concerning companies' efforts to combat online child sexual exploitation and abuse (CSEA). While high-level and lacking the detail of the OECD framework, it indicates a consensus regarding the need for consistency and global standards as a trust measure.

Question 23: Should the Code have a transition period or transition periods for specific issues? Which areas touched on in this Call for Inputs may VSPS providers require time to transition the most? What time frame would be reasonable for a transition period?

Transition periods and timelines should align with equivalent measures at the EU level and in conjunction with other regulatory interventions and scaled according to the size and capacity of the providers with priority to the introduction of requirements for the largest platforms – for whom obligations under DSA already apply.

⁵³ <https://www.oecd.org/digital/transparency-reporting-on-terrorist-and-violent-extremist-content-online-8af4ab29-en.htm>

⁵⁴ See <https://www.oecd.org/digital/children-digital-environment/>

Call For Inputs: Online Safety

Department of Children, Equality, Disability, Integration and Youth

The Department of Children, Equality, Disability, Integration and Youth (DCEDIY) welcomes the significant progress to develop Ireland's first binding online safety code for video-sharing platform services.

4. Overall Approach to the Online Safety Code

The new online safety code should be strongly informed by children's rights, particularly as set out in the UN Convention on the Rights of the Child, which Ireland ratified in 1992. As it applies to children and young people up to the age of 18, the code should reference and be grounded in specific articles of the UNCRC, including Article 12 on the right of children to express their views and be heard, and Article 13 on their right to free expression. Realising Article 17, which recognises "the important function performed by the mass media and ... access to information and material from a diversity of national and international sources, especially those aimed at the promotion of his or her social, spiritual and moral well-being and physical and mental health" should be central to the mission and operation of the code. The Online Safety Code should clearly outline the child rights function it is fulfilling, and those who operate the code should be aware of all of our responsibilities to uphold children's rights.

A Child Rights Impact Assessment should be carried out as part of the implementation of the Online Safety Code overall, looking specifically at the likely impacts it will have on children. Furthermore, as part of compliance with the code, VSPS should be required to carry out regular Child Rights Impact Assessments on the extent to which their content upholds and promotes the inalienable rights of children and young people.

5.1 Online Safety Features for Users

The robust proposals for online safety features for users are welcome. When VSPS providers design and implement online safety features for their platforms, they should ensure they are all fully available in a child-friendly format, so that all internet users, whatever their age, can report content in violation of the code.

5.1.3 Age Verification and Age Assurance

Age Verification and Age Assurance is welcome, and the document already outlines many of the complexities of balancing robust age verification with the possible collection of sensitive data. However, another concern with robust age verification systems such as the need to provide identity documents or use a proposed European Digital Identity is the potential for digital exclusion of young people from marginalised groups. Some vulnerable groups in society may be less likely to have identity documents than others, and it will be important to study what effect such a system would have on digital access for young people who are of age to access content, but don't have the required ID. Scaling the robustness of age verification with the potential harm of content may mitigate some of these issues.

5.1.4 Content Rating Feature

Content rating would be a broadly welcome feature, and it is good to allow parents or guardians to have an insight into the type of content which their child might access. However, it is important that such a rating system also gives the child or young person themselves an easy-to-understand rating for the content they will view. Content rating should be designed to be child-friendly, so children and young people themselves can also make informed decisions about what they can access.

5.1.5 Parental Controls

In addition to blocking harmful content and privacy settings, the ability to limit a child's use of service is important. It is important to distinguish the issues of children's access to potentially harmful content, age-inappropriate content and other potential harms that can result from spending excessive time viewing content which is not, in itself, harmful. While the problems of exposing children to harmful content are well documented, research has also demonstrated the specific harms such as poor mental health and sleep deprivation from overuse of online platforms. Recent research [reported on in the Irish Times](#) showed correlations between poorer mental health and a lack of sleep, and linked the lack of sleep to smartphone use.

We would welcome parental controls by default and clear and accessible guidance to parents on how to set these controls.

5.1.6 Media Literacy

This should be provided in child friendly format. The developers of media literacy tools should develop them with the key concerns of children and young people in mind, studying available data on issues such as body image.

5.3.5 Harmful feeds and recommender systems

It is vital that aggregate content which could cause harm to a child or young person is interrupted/mitigated. 5.3 Possible Additional Measures and Other Matters.

5.3.3 Safety by Design

We would be supportive of the proposal to require VSPS providers to follow a 'safety by design' approach when they introduce new features. We note that one approach to reflecting this in the Code would be to require VSPS providers to publish a 'Safety by Design' statement setting out how they consider online safety when developing or enhancing services. We would support the proposed requirement to prepare a "Safety Impact Assessment" whenever services are being developed or enhanced, with sign-off of the risk assessment and proposed mitigation measures by an executive staff member of the VSPS provider with appropriate experience and responsibilities.

This proposal would seem to be in alignment with the obligations contained in the Children First Act 2015, which requires organisations providing 'relevant services' to children to keep children safe

from harm while they are using the service, to undertake a risk assessment and to develop a Child Safeguarding Statement (CSS) setting out the procedures in place to manage any risk identified. These should include policies and procedures on child safeguarding awareness and training and on the reporting of child protection concerns.

The types of organisations to which these statutory obligations apply are set out in Schedule 1 to the Act. They include any work or activity which consists of the provision of educational, cultural, recreational or leisure or social activities to children. Note the onus is on organisations to examine the legislation to determine whether any aspect of their work brings them within the definition of 'relevant services'.

Furthermore, as part of the Action Plan for Online Safety, in January 2019 the (then) Department of Children and Youth Affairs published an addendum to the Children First National Guidance to include a specific reference to the need to consider online safety in the preparation of risk assessments and Child Safeguarding Statements. The addendum is available on the following link <https://www.gov.ie/en/publication/c7ee34-action-plan-for-online-safety/>.

More background information about Children First can be found in the Appendix.

Annex 1

Clarification is requested on the exact make-up of the proposed Youth Advisory Committee. The initial reference to this committee (p. 5) states that it will seek representation from young people who are 25 years of age or younger, or of not more than 25 years of age. However, when the committee is referenced in the Annex (p.26), this refers to half of the members being under the age of 25. The Department would welcome a focus on those aged 0-24, to align with the forthcoming policy framework for children and young people.

This committee will also draw membership from those working with children and young people, which is welcome. However, there does not appear to be a clear rationale for limiting this section of the membership to only over-25s. In practice, most people working with and for children and young people will be over 25, but disqualifying younger professionals has no rationale in the text and would not be supported by DCEDIY. Overall, the Youth Advisory Committee structure is very welcome and an excellent step to ensuring children and young people's voices are heard on this important topic.

Appendix

Background Information about Children First

Tusla, the Child and Family Agency has a statutory duty under the Child Care Act 1991 to promote the welfare of children who are not receiving adequate care and protection. In doing so, it relies heavily on individuals reporting concerns about children, in accordance with *Children First: National Guidance for the Protection and Welfare of Children 2017* and the Children First Act 2015.

The Children First Act 2015 which was fully commenced in December 2017 provides for a number of key child protection measures, including raising awareness of child abuse and neglect, providing for reporting and management of child protection concerns and improving child protection arrangements in organisations providing services to children. The Act operates side-by-side with the non-statutory obligations provided for in *Children First: National Guidance for the Protection and Welfare of Children*.

The Guidance has been in place since 1999 and was fully revised in 2017 to include reference to the provisions of the Act. The Guidance sets out definitions of abuse, and signs for its recognition. It explains how reports about reasonable concerns of child abuse or neglect should be made by the general public and professionals to Tusla. It also sets out safeguarding best practice to assist any organisation providing a service to children to create a safe environment. The Children First Act and Guidance are intended to empower service providers, members of the public and all people working with and caring for children to recognise and confront suspected child abuse.

The Children First Act, provides for a number of key child protection measures and can be best summarised as having three key elements. The first is that the Act provides for mandatory reporting of child protection concerns by key professionals, including teachers, gardaí and health care professionals. Under the Act, mandated persons are required to report child protection concerns at or above a defined threshold to Tusla. Mandated persons are people who have contact with children and families and who, because of their qualifications, training or employment role, are in a key position to help protect children from harm. The list of mandated persons is set out in Schedule 2 of the Act.

The second key element is that the Children First Act places specific obligations on particular organisations that provide 'relevant services' to children and young people, including a requirement to keep children safe from harm while they are using the service, to carry out a risk assessment and to develop a Child Safeguarding Statement. This is a written statement that sets out the service provided and the principles and procedures in place to ensure, as much as possible, that a child or young person using the service is safe from harm.

The third key element of the Children First Act was establishing the Children First Inter-Departmental Implementation Group, on which each Government Department, Tusla, the HSE and An Garda Síochána is represented, on a statutory footing. The functions of the Implementation Group include promoting compliance by Government Departments with their obligations under the Act. The Group also provides a forum for members to raise child welfare and protection issues of general concern, or with a cross-departmental or cross-sectoral dimension across the various sectors.

Overall, the Children First Act represents an important addition to the legal framework for child protection in Ireland and it helps to ensure that child protection concerns are brought to the attention of Tusla without delay.

More information about Children First can be found on the DCEDIY website
<https://www.gov.ie/en/policy-information/d1b594-children-first/>

Submission from

Ministry of the Interior and Kingdom Relations, Netherlands

Question 1: What do you think our main priorities and objectives should be in the first binding Online Safety Code for VSPS? What are the main online harms you would like to see it address and why?

In the Dutch legislation we choose to follow the categorisation laid down in AVMSD (content which may impair the physical, mental or moral development of minors, content that incites violence or hatred against the listed groups, content which constitutes a criminal offence under EU law, and certain commercial communications as listed in the directive). We are interested in the broader approach that is taken in Ireland's 2009 Act and that will possibly be copied into the Online Safety Code for Video-Sharing Platform Services.

Especially concerning the protection of children we think it is important that not only the content itself is taken into account, but also the way the content is offered to the user. For example, one video about a diet could be harmless, but it can become harmful if the user ends up in an information rabbit hole about diets offered by algorithms. This could also be a topic to be addressed in the code.

Question 2: What types of online harms do you think should attract the most stringent risk mitigation measures by VSPS? How could we evaluate the impact of different types of harms e.g. severity, speed at which harm may be caused? Is there a way of classifying harmful content that you consider it would be useful for us to use?

NL is currently developing a Children's rights impact assessment. (KIA) This instrument is being developed to map the risks for children's rights online. Based on this KIA, an estimate can be made of the possible risks associated with the use of a certain digital service. Once these risks have been identified, an assessment should then be made by the provider of that online service of which measures are most effective to mitigate these risks. When the development of the children's rights impact assessment is finished, we would be of course willing to exchange thoughts on this topic.

Question 6: How should we design the Code to minimise the potential for conflict and maximise the potential for synergies in how platforms comply with it and the DSA?

The NL does not have the answer to this question. We however are very interested in Ireland's approach and the relationship between its national legislation and the DSA. As the DSA is maximum harmonization, it is interested to us in what way national legislation concerning similar topics and the DSA can coexist.

Question 8: How should we ask VSPS providers to introduce a feature that allows users to declare when videos contain advertising or other type of commercial communications? Should the Code include specific requirements about the form in which the declaration should take? What current examples are there that you regard as best practice? Dutch influencers with over 500.000 followers are required to be transparent about advertising in their videos. They have to make this known by declaring that the content contains advertising by indication 'advertising', 'advertisement', 'paid promotion', or '#ad', when posting a video. They can also make use of the options offered by certain VSPS to designate a video as 'advertising'.

Question 9: How should we ask VSPS providers to introduce and design a flagging mechanism in the Code? How can we ensure that VSPS providers introduce the mechanism in a user-friendly and transparent way? How should we ask VSP Providers to report the decisions they've made on content after it has been flagged? To what extent should we align the Code with similar provisions on flagging in the DSA?

The NL supports an adequate flagging mechanism as well as transparency on the mechanism and decision-making process. In that regard, we favor alignment with the provisions in the DSA, such as a trusted flagger status. However, the possibility of establishing a flagger system not exclusively for reporting illegal content as the DSA requires in Article 22 could also be looked at. It would be useful to see if harmful content could also be included in a trusted flagger system.

In addition to this trusted flagger mechanism as described in the DSA, we are currently looking into the possibility of setting up a low-threshold hotline in the Netherlands where Dutch citizens can ask for help with removal requests for online content or other content-related questions. To this end, we are currently in discussions with the private sector including hosting service providers, social media platforms and civil society. We expect to start a pilot in October or November of 2023, which will run until April or May of 2024. We expect to have more knowledge on this issue after the pilot.

Question 10: What requirements should the Code include about age verification and age assurance? What sort of content should be shown by default to users who are logged out or in private browsing mode and whose age cannot be verified or assured? What evidence is there about the effectiveness of age estimation techniques? What current practices do you regard as best practice? Where accounts are not age verified should default privacy settings be used, should content default to universal content and should contact by others be more limited? The NL supports adequate age verification systems that cannot be easily circumvented and we are currently working on a list of minimum requirements for adequate age verification systems (privacy protection, inclusion, security and system robustness requirements) and an assessment framework in which requirements with regard to age verification and the appropriate application of this are mapped out for each risk category. We see that the fitting type of age verification is very dependent on the specific context, this framework is designed to assess per risk category what the most fitting age verification/assessment method is. We expect to finish the development of the first version of this framework this fall.

Question 15: How should we ask VSPS providers to address content moderation in the Code? Are there any current practices which you consider to be best practice? How should we address automated content detection and moderation in the Code?

When it comes to content moderation, it remains crucial that moderators possess a deep understanding of both the linguistic nuances and cultural context of the content at hand. To achieve this, providers of VSPS must have access to a diverse workforce. This measure serves to mitigate the potential for misinterpretation or mistranslation, which can subsequently result in wrongful decisions. Besides making sure content gets reviewed quickly, human oversight in content moderation is important. In instances where human oversight is unfeasible by default, it is incumbent to transparently communicate to users that their content underwent assessment via an automated system. This level of transparency empowers users by providing insight into the moderation process, the decisions arrived at, and avenues to file a complaint if they feel that an incorrect assessment has taken place.

Next to this, we are also mapping the moderation policies of social media platforms and which practices are currently used. When completed, we will be happy to discuss this with you.

Question 18: What approach do you think the Code should take to risk assessments and safety by design? Are there any examples you can point us towards which you consider to be best practice?

NL is developing various tools to better protect children online. These instruments point out to providers of online services and products in various phases (both during development of online products or services and when these services and products are already being offered) the children's rights that must be taken into account. For example, the existing [Online Children's Rights Code](#) will be updated and transformed into a more practical tool for designers of digital services and products. As mentioned in the answer to question two, a children's rights impact assessment is also being developed, in which the risks of an online service or product for children's rights are mapped out. Furthermore the University of Utrecht has developed on behalf of the Dutch government the Fundamental Rights and Algorithm Impact Assessment ([FRAIA](#), Dutch: IAMA). The FRAIA helps to map the risks to human rights in the use of algorithms and to take measures to address these risks. FRAIA creates a dialogue between professionals who are working on the development or deployment of an algorithmic system. The client is responsible for the

implementation of the FRAIA. This results in addressing all relevant points for attention when using algorithms in a timely and structured manner. This prevents organizations from using algorithms of which the consequences are not yet clear. The FRAIA also reduces the risks of carelessness, ineffectiveness, or infringements of citizens' rights.

NL has a 'by-design' approach in mind, in which specifically children's rights should be taken into account from the very beginning of the development of new products or online services and which should reoccur throughout the life cycle of an online service or product.

Question 20: What approach do you think we should take in the Code to address feeds which cause harm because of the aggregate impact of the content they provide access to? Are there current practices which you consider to be best practice in this regard? The risk for users to end up in an information rabbit hole also has NL's attention. The DSA sets conditions for recommendation algorithms. VLOPs and VLOSEs should now offer their users at least one option that allows them to use the service without that service using profiling for making recommendations. In addition, all online platforms – regardless of the number of users – are required to be transparent about the main parameters used in their recommendation systems. They must also be transparent about any options for users of the service to change or influence these parameters. When users have the ability to customize the recommendation system that functionality should be easily accessible. We hope and expect that these regulations will help protect (minor) users against the harm caused by the aggregate impact of content. However, we would be interested in seeing if Ireland is planning to implement any additional requirements regarding this topic.

Comisiún na Meán Call for Inputs: Online Safety

Developing Ireland's First Binding Online Safety Code for Video-Sharing Platform Service

Introduction

NewsBrands Ireland and Local Ireland welcome the opportunity to respond to the above call for inputs and set out here our high-level comments as they relate to content published via on-demand services by news publishers.

NewsBrands Ireland is the representative body for national news publishers. We promote the vital contribution made by our members' trusted journalism to society and democracy and highlight the importance of a fair and balanced legislative framework that supports public service journalism.

Local Ireland is the promotional brand of the Regional Newspapers and Printers Association of Ireland, formerly the Provincial Newspaper Association, founded in 1919, and the oldest newspaper association in Ireland. Regional news publishers in print and online are vital to the communities they serve. No other media can consistently deliver high quality, professional content at such a hyperlocal level.

Role of Journalism in Democracy

News publishers are pillars of democracy, providing crucial information, insights and perspective to citizens on the events shaping our society. The journalism produced by the member news publishers of NewsBrands and Local Ireland across print and digital platforms is read by 4 out of 5 adults, 82% of the population. It helps to inform citizens with public interest news and information and contributes to the democratic debate. It is imperative that our members' right to right to freedom of expression guaranteed by the Irish Constitution and the European Convention on Human Rights is not undermined.

As such, we argued previously in earlier submissions on the OSMR that there should be an exemption for publisher content, which is already regulated through the Press Council and subject to strict defamation laws.

Online Harm

We recognise and support the need to mitigate online harm, especially of minors and young persons. We are founding members of the Press Council of Ireland and the print and digital platforms represented by both organisations subscribe to the Press Council Code of Practice. The Press Council Code encompasses many of the areas that will be covered by the Online Safety Code, including the protection of minors and the prohibition of material encouraging or promoting self-harm, suicide and eating disorders.

Further, our members are currently subject to much stricter legal criteria than apply to most tech companies, even where the content in question is generated by third parties by way of, for example, commentary below an online report.

We also recognise that in a complex news landscape, media literacy is crucial. It means more than identifying ‘fake news’; it is about understanding journalistic processes and their value, how news is presented online and how it is regulated. Irish news publishers recognise the vital importance of news and media literacy to democracy. We are an active member of Media Literacy Ireland and we run a free news literacy and student journalism programme for secondary schools. [Press Pass](#), which has been completed by over 110,000 transition year students to date, is designed to empower students to recognise responsibly produced news and learn how to produce their own journalism.

VSPS Moderation

We are concerned that our members’ already regulated and trusted journalism, disseminated via an on-demand service, will become subject to policing by tech companies, and their interpretation when seeking to fulfil their duties and responsibilities.

Consequently, material published in the public interest could be blocked by tech companies through their operation of compliance systems which are likely to rely heavily upon algorithms of necessarily limited sensitivity and the increasing use of AI. There is a high risk that decisions could be taken without consideration of the context of the many contentious issues that are covered by news publishers as part of their role to inform and educate citizens.

Further, there is a risk that if the material that has been removed has been generated by a third party by way of, say, an opinion piece, a quote for an article or a below the line comment, our members could be sued by the affected party despite having had no input into the decision complained of.

News publisher content which touches on defined harmful online content categories risk being taken down or down-rated via algorithm by overly zealous moderation activity by platforms. The difficulties that this could give rise to are exacerbated by the fact that the

terms and conditions of service of most tech companies give little or no redress to affected parties when material is removed or edited by them.

Conclusion

We recognise the key role of Comisiun na Meán in developing Ireland's first binding online safety code for video-sharing platform services. It is important however that careful consideration is given to public service journalism produced by trusted news publishers that is stored on these platforms and recognition is given to its essential role in society. As stated previously, our original submissions on the OSMR sought an exemption for news publisher content that is already subject to strict defamation laws and regulated through the Office of Press Ombudsman and Press Council of Ireland.

Ends/ 4 September 2023

NewsBrands Ireland
www.newsbrands.ie
Ann Marie Lenihan, CEO
[REDACTED]

Local Ireland
www.localireland.info
Bob Hughes, Executive Director
[REDACTED]

September 2023

NWC Call For Inputs: Online Safety

Introduction

Founded in 1973, the National Women's Council (NWC) is the leading national women's membership organisation in Ireland. NWC represents and derives our mandate from our membership, which includes over 190 groups and organisations from a diversity of backgrounds, sectors and locations across Ireland. Our mission is to lead and to be a catalyst for change in the achievement of equality for women. Our vision is of an Ireland and of a world where women can achieve their full potential and there is full equality for women.

NWC chairs the National Observatory on Violence Against Women an independent network of over 24 grassroots and national organisations that convene quarterly to monitor progress on violence against women in Ireland. NWC established and chaired the National Advisory Committee supporting the Dept. of Higher Education's Framework Safe, Respectful, Supportive and Positive – Ending Sexual Violence and Harassment in Irish Higher Education Institutions, the work of this Committee is now mainstreamed into the Higher Education Authority.

NWC has welcomed the Online Safety and Media Regulation Act 2022 and the establishment of an Online Safety Commissioner to oversee the new regulatory framework. NWC also welcomes the opportunity to input in relation to the online safety codes for video-sharing platform services, especially the consultation's emphasis on the protection of children and the general public from online harms while upholding and promoting human rights, by requiring VSPS providers to introduce online safety features for their users and to moderate content more effectively. It will be crucial that this Code aligns with the Third National Strategy on Domestic Sexual Gender-Based Violence, as part of a whole of government approach to end violence against women and girls. Ireland is in a unique position as the EU HQ of many leading technology companies and video-sharing platform providers to contribute to making Europe safer for women and girls.

Alarming, 1 in 2 women who had suffered intimate relationship abuse experienced abuse online using digital technology¹. Cyber violence against women is an increasing problem worldwide (including cyberstalking, image-based sexual abuse, gender-based slurs and harassment, 'slut-shaming',

¹ https://www.toointoyou.ie/app/uploads/2022/10/one_in_five_women_report_womens_aid_2020.pdf

pornography, 'sextortion', rape and death threats, 'doxing', and electronically enabled trafficking). Furthermore, the Department of Justice (2020)² Report on the Public Consultation - Hate Speech and Hate Crime in Ireland shows that there is considerable disquiet at the use of media and online platforms by public figures to promote racist stereotypes and harmful myths in order to generate attention for their campaigns and that much of social media prejudice is expressed against Travellers. BelongTo revealed that a shocking 87% of LGBTQ+ young people have seen or experienced anti-LGBTQ+ hate and harassment on social media in the past year³. Actions must be taken against anti-LGBTQ+ content, as only 21% of LGBTQ+ youth who reported abusive or harmful LGBTQ+ content saw action from a social media platform⁴.

Moreover, NWC has strongly stressed the importance of the inclusion of the harms of pornography in the SPHE⁵ curricula and the Third National Strategy on DSGBV to address this issue⁶, and in its National Observatory Shadow Report to Grevio (2022)⁷. NWC, through the Beyond Exploitation Campaign (2020)⁸ has highlighted the harms of pornography on children and young people, by influencing expectations, normalising sexual behaviour based on misogynistic, and often abusive and violent, models of sexual expectations. Pornography also has an impact on gender equality and is a form of sexual exploitation and violence against women.

Extensive research is now available about the harms of pornography on children and young people⁹, including how hardcore, explicit porn is widely available. 1 in 3 children say they've seen explicit, hardcore porn by age 12 and its misleading, degrading, and objectifying impact is profound (as it shapes children and young people's perception of consent, sexual violence, gender equality, sexuality, and intimate relationships at an extremely young age and without any context). In the UK, 44% of males ages 11–16 who saw hardcore porn said it gave them ideas about the type of sex they wanted to try¹⁰.

² Department of Justice (2020). Legislating for Hate Speech and Hate Crime in Ireland Report on the Public Consultation 2020. Available at <https://assets.gov.ie/237922/07cb2005-2712-4808-9b48-348f224806b5.pdf>

³ <https://www.belongto.org/87-of-lgbtq-youth-report-hate-and-harassment-online/>

⁴ <https://www.belongto.org/87-of-lgbtq-youth-report-hate-and-harassment-online/>

⁵ https://www.nwci.ie/images/uploads/NWC_Submission_on_Senior_Cycle_SPHE_Redevelopment.pdf

⁶ Department of Justice (2022), Implementation Plan - Zero Tolerance Third Domestic, Sexual and Gender-Based Violence, Action 1.5 and actions 1.1.4 1.3.7

⁷ National Observatory on Violence against Women and Girls, Shadow Report to GREVIO in respect of Ireland (2022)

⁸ Beyond Exploitation (2020), Submission to Third National Strategy on Domestic, Sexual and Gender-Based Violence

⁹ <https://www.culturereframed.org/the-porn-crisis/>

¹⁰ https://www.mdx.ac.uk/_data/assets/pdf_file/0021/223266/MDX-NSPCC-OCC-pornography-report.pdf

In this call for inputs, NWC will focus on the Online Harms and the issues that the Online Safety Code should address, in particular on the harms of pornography and the need to tackle it as a priority for children and young people, as well as for gender equality in general.

Question 1: What do you think our main priorities and objectives should be in the first binding Online Safety Code for VSPS? What are the main online harms you would like to see it address and why

NWC's focus in this submission is on combatting harms to women and girls in line with the current whole of government Zero Tolerance strategy on Domestic, Sexual and Gender-Based Violence, and we urge that the main priorities and objectives for the first binding Online Safety Code for VSPS should include the following specific area:

Combatting exposure of children to pornography, particularly in relation to the suspected link between such exposure and the increase in harmful sexual behaviour among children and young people

1. Pornography and the increase in harmful sexual behaviour among children and young people

Regular consumption of pornography is unfortunately commonplace among children, it has become normalised to the extent that in the words of the UK's Children Commissioner, children 'cannot opt-out'.¹¹ Inappropriate exposure to pornography (most of which is extreme, violent, and degrading to women) at a very young age is a complex, multi-faceted social problem. This inescapable digital environment of misogyny and brutality is where children and young people spend much of their online lives and it cannot but be a contributing factor to harmful attitudes to sex, relationships and gender, many believe it to be a driver of sexual violence.¹² The biggest determiner of when children first view pornography is the age at which they get a phone, and research by Irish charity *Cybersafe Kids* found that 95% of 8–12-year-olds owned their own smart device (an increase of 2% on the previous year's figure), with 87% of 8–12 year-olds having their own social media and/or instant messaging account (an increase of 3% on 2021)¹³.

Sexual Assault Treatment Units (SATUs) in Ireland have seen a rise in the number of victims of peer-to-peer violence they treat. The SATU annual report for 2022 shows that 20.1% of victims attending their centres were 18 years or below.¹⁴ The SATU network provides care to anyone over the age of 14 who has been sexually assaulted. On occasion, the SATU service may care for children under the age of 14, when

¹¹ 'A Lot Of It Is Actually Just Abuse'- Young People And Pornography | Children's Commissioner For England (no date) Children's Commissioner For England. Available at: <https://www.childrenscommissioner.gov.uk/resource/a-lot-of-it-is-actually-just-abuse-young-people-and-pornography/> (Accessed: 30 August 2023).

¹² Lally, C. (2018) 'Is pornography driving increased sexual violence in Ireland?', *The Irish Times*, 26 May. Available at: <https://www.irishtimes.com/news/crime-and-law/is-pornography-driving-increased-sexual-violence-in-ireland-1.3508313> (Accessed: 31 August 2023).

¹³ Cybersafe Kids (2023) *ACADEMIC YEAR IN REVIEW 2021–2022*. Cybersafe Kids. Available at: https://www.cybersafekids.ie/wp-content/uploads/2022/09/CSK_YearInReview_2021-2022_FINAL.pdf (Accessed: 30 August 2023).

¹⁴ National Women and Infants Health Programme (2023) *SATU ANNUAL REPORT 2022*. HSE. Available at: <https://www.hse.ie/eng/services/list/5/sexhealth/sexual-assault-treatment-units-resources-for-healthcare-professionals/satu-2022-annual-report.pdf> (Accessed: 23 August 2023).

paediatric services are unavailable and there is an acute forensic need for attendance. The data presented in the 2022 report shows that adolescents are significantly represented in SATU attendances every year. Furthermore, between 2017 and 2022 SATUs noted an increase of 37% in the number of adolescents presenting to the network and refer to it as a 'key emerging theme'. Gardaí recorded 97 suspected sexual assault and rape offenders aged under 16 in 2019, and 79 in 2020.¹⁵ If the extremely low reporting rates of such assaults are taken into account this is a sizeable number. The number of children under the age of 18 alleged to have abused other children the same age or younger increased by 18% in 2022, according to figures from Rape Crisis Network Ireland (RCNI).¹⁶ The CSO report that in one in seven (15%) cases of detected sexual violence in 2020, both the victim and suspected offender were under 18 years of age.¹⁷ During the same period, Donegal Rape Crisis Centre had seen a 50% increase in victims under 16 seeking help with the youngest just 12 years old and they have also noted that the level of physical violence accompanying sexual violence has seriously escalated.¹⁸

This chimes with the situation in the UK, where the Children's Commissioner commissioned research on an apparent similar surge in peer-to-peer sexual violence with aggravated physical violence using an innovative methodology to explore the link between exposure to pornography and sexual violence.¹⁹ The research examined transcripts of interviews between medical personnel or police and children who were victims or perpetrators of sexual assault/abuse between 2012 and 2022. The results showed that in 50% of the cases the transcripts contained references to acts of sexual violence commonly portrayed in porn. The most frequently occurring categories of physical aggression were strangulation, choking or slapping, with name-calling also prevalent. An additional review of some of these cases found children themselves suggesting direct links between pornography exposure and the harmful sexual behaviour exhibited.

The really interesting part of this research is the longitudinal analysis which shows that a minority of the police transcripts (10%) overall mentioned pornography, although that had risen to nearly a quarter of cases between 2017 and 2022. This timeframe coincides with the growth in access to smart devices by adolescents and teens. The references were most often to: watching pornography; girls being seen as a porn star; specific types of porn; or porn sites. This is compelling evidence that abusive acts represented

¹⁵ Edwards, R. (2022) 'Concern at rise in number of children alleged to have sexually abused the young', *The Independent*, 16 October. Available at: <https://www.independent.ie/irish-news/concern-at-rise-in-number-of-children-alleged-to-have-sexually-abused-the-young/42069817.html> (Accessed: 31 August 2023).

¹⁶ Edwards, R. (2022) *ibid*

¹⁷ Central Statistics Office (no date) *Press Statement Recorded Crime Victims 2021 And Suspected Offenders 2020 - CSO - Central Statistics Office*. Available at: <https://www.cso.ie/en/csolatestnews/pressreleases/2022pressreleases/pressstatementrecordedcrimevictims2021andsuspectedoffenders2020/> (Accessed: 31 August 2023).

¹⁸ Edwards, R. (2022) *ibid*

¹⁹ Children's Commissioner (2023) *Evidence on pornography's influence on harmful sexual behaviour among children*. Gov.UK. Available at: <https://assets.childrenscommissioner.gov.uk/wpuploads/2023/05/Evidence-on-pornographys-influence-on-harmful-sexual-behaviour-among-children.pdf> (Accessed: 23 August 2023).

in pornography are occurring in sexual assaults and violence against girls. It's vital that similar research is conducted here in Ireland, replicating the UK methodology to shed more light on the link between the increase in the numbers of children presenting as victims and perpetrators of sexual violence, often with accompanying aggravated physical aggression, and exposure to pornography. The horrific sexual assault and murder of Ana Kriégel in Dublin in 2018 is a tragic example of an outcome of disordered views on sexuality and gender as a result of repeated exposure to violent pornographic images among young boys. The details that emerged about the online life of one of the 13-year-olds convicted of the murder were very troubling. Boy A had 12,500 images on two devices that gardaí found in his bedroom, the vast majority of which were of a pornographic and brutal nature.²⁰

2. Problematic issues with specific VSPs and children's exposure to pornography

VSPs are where most children and young people first encounter porn, with the social media site X formerly known as *Twitter* the online platform where they report they were most likely to have initially encountered it²¹ (this is certainly the case in the UK, there's no disaggregated age information available on social media use in Ireland, but we do know that Pornhub is very frequently used here, and in fact was the 12th most visited website in Ireland in July 2023²²). Ofcom's *Children's and Parents' Media Literacy 2021* study showed that use of video-sharing platforms was the most-cited online activity among all children aged 3-15 (94%). Use among children increases with age, with almost all (98%) 12-15-year-olds reporting they watch content on video-sharing platforms.²³ Amongst older teenagers, 15-17, Twitter is ranked number 4 in the top twenty video-sharing platforms by reach, with an online reach of 62%.²⁴ Many people encounter pornography on *Twitter* through the accounts of content creators using the platform to drive traffic to their *Onlyfans* page. As *Onlyfans* doesn't allow advertising, content creators there use other VSPs to promote their sites.

X/ *Twitter* settings are organised in a particular way on these sexualised accounts to avoid being banned, the account owners fulfil the platform's requirement of tagging the profile as 'sensitive' (these accounts are commonly indicated on the platform in bios as NSFW- not safe for work). This can be done very simply within privacy/security setting to indicate that the posts from the account on the platform may contain 'sensitive' material. The NSFW search term can then be used as a way of accessing porn on the platform.

²⁰ Gallagher, C. (2019) 'Ana Kriégel murder trial: jury not told of porn found on Boy A's phone', *The Irish Times*, 1 January. Available at: <https://www.irishtimes.com/news/ireland/irish-news/ana-kriegel-murder-trial-jury-not-told-of-porn-found-on-boy-a-s-phone-1.3929624> (Accessed: 31 August 2023).

²¹ Children's Commissioner (2023b) 'A lot of it is actually just abuse'- Young people and pornography. Gov.UK. Available at: <https://assets.childrenscommissioner.gov.uk/wpuploads/2023/02/cc-a-lot-of-it-is-actually-just-abuse-young-people-and-pornography-updated.pdf> (Accessed: 31 August 2023).

²² Most visited websites Ireland July 2023 (no date). Available at: <https://www.similarweb.com/top-websites/ireland/#:~:text=google.com%20is%20ranked%20%231,in%20Ireland%20is%20independent.ie>. (Accessed: 31 August 2023).

²³ Ofcom (2022) *The VSP Landscape Understanding the video-sharing platform industry in the UK*. Ofcom. Available at: https://www.ofcom.org.uk/data/assets/pdf_file/0030/245577/2022-vsp-landscape.pdf (Accessed: 31 August 2023).

²⁴ Ofcom (2022) *ibid*

These initial searches can lead children to other VSPS due to cross-posting on multiple accounts across *Instagram*, *Facebook* and *TikTok* in particular. While children or young people will not necessarily be successful in becoming *Onlyfans* consumers as their age verification procedures are relatively robust, they will simply by virtue of coming across *Onlyfans* ‘teaser’ videos and interacting with them, be fed more explicit content in their ‘timeline’ and seek more. The platforms may also lead the user to *Pornhub* as many content creators have accounts on that VSPS too (through their ‘model program’). However, it’s also simply, and very frequently the case that the child or young person uses hashtags to search, for example #porn #hardcoreporn #onlyfans which generates many results in *X/Twitter* and once searched, the algorithm will then populate the timeline with age-inappropriate and pornographic content. Many information sources are available to the *Onlyfans* creator helping them to use *X/Twitter* to boost the visitors to their camming site. One such resource points out

‘many of the popular social media platforms restrict NSFW content or adult content, Twitter is very lenient in this regard. You can post any type of content for the promotion of your fan page. That’s the primary reason for choosing Twitter to promote and advertise your content’.²⁵

Tackling just this one issue on *X/Twitter* which facilitates the masking of pornographic material, would reduce significantly children’s and young people’s exposure to porn on that platform where most indicate they first encounter it. Also removing ‘porn’ as a searchable item would help, this has been done with *Instagram* and while there is hardcore pornography still on the VSPS it is less pervasive and less easy to find.

TikTok is listed number 6 in the top twenty video-sharing platforms by reach, with an online reach of 66% among 15 to 17 year olds in the UK.²⁶ Thirteen is the age at which children can officially open *TikTok* accounts although in practice children far younger use the VSPS. Again here, *TikTok* is used by *Onlyfans* content creators and others in the sex industry on *Pornhub* for example, to promote their explicit content without getting their videos removed, by using specific filters on their images, for example by modifying their adult images to look like paintings.²⁷ Another way of doing this is by using an explicit image or video as a profile picture which also circumvents moderation rules.²⁸

²⁵ *How To Promote the Onlyfans Page on Twitter?* (2022) *Medium*. Available at: <https://medium.com/betteronlyfans/how-to-promote-the-onlyfans-page-on-twitter-7d4451aa48f4> (Accessed: 31 August 2023).

²⁶ Ofcom (2022) *ibid*

²⁷ *Some OnlyFans Creators Have Found A Loophole To Put Their Nudes On TikTok* (2022) *NBC News*. Available at: <https://www.nbcnews.com/pop-culture/viral/onlyfans-creators-loophole-nudes-tiktok-ai-filter-rcna56484> (Accessed: 31 August 2023).

²⁸ *TikTok Loophole Sees Users Post Pornographic And Violent Videos* (2021) *BBC News*. Available at: <https://www.bbc.com/news/technology-56821882> (Accessed: 31 August 2023).

Similar to the cross-fertilisation of porn between *X/Twitter* and *Onlyfans* and *Pornhub*, *TikTok's* superior editing tools and an increasing demand for porn in a *TikTok* style on *Pornhub* is driving the rate at which *TikTok* videos are posted to other VSPS and this solidifies the connection between *Pornhub*, *Onlyfans* and *TikTok* as platforms that filter users towards each other. While nudity and sexual activity are not allowed according to *TikTok's* rules, very many explicit videos get past the moderation system. Sex industry performers use *TikTok* video production tools to create a video, then screen record it and upload it directly to *Onlyfans* without ever posting to *TikTok*. *TikTok* is effectively operating as a marketing tool for some content creators in the sex industry.²⁹ Issues relating to VSPS use by the Tate brothers in their alleged sex-trafficking operation and enforced sexual exploitation of women on *TikTok* and *Onlyfans* is likely to be admitted as evidence in the case being taken by the Romanian Government.³⁰ While most *Onlyfans* content is of a sexual nature, much of it is not in the extreme category, though there is also a quantity of hardcore, degrading and deeply misogynistic material on it. The issue is primarily the filtering of consumers of porn from one VSPs to another in particular those like *Pornhub* and *Redtube* which are free and contain a really dizzying array of extremely violent and brutal pornography.

TikTok's use for the production and dissemination of Child Sexual Abuse Material (CSAM) is also widely acknowledged in the wake of a *Forbes* investigation into this issue.³¹ This distribution of pornographic materials involved the sharing of passwords to single accounts so multiple users could log in privately (anonymously) and predators could 'meet' there and share images of minors. *Forbes* found that there was a huge volume of these post-in-private accounts and that new ones popped up as quickly as they were banned or shut down. While this specific investigation related to CSAM there's no reason to believe that this strategy is not also being used for other kinds of pornography

Question 3: Do you have reports, academic studies or other relevant independent research that would support your views? If you do, please share them with us with links to relevant reports, studies or research.

Some relevant links are mentioned in the introduction and in response to question 1 above. In this section we would like to highlight other relevant research that support our concerns and recommendations.

²⁹ Schofield, D. (2021) *TikTok Has Accidentally Conquered The Porn Industry*, *WIRED UK*. Available at: <https://www.wired.co.uk/article/tiktok-nsfw> (Accessed: 31 August 2023).

³⁰ *Andrew Tate Prosecution Files Reveal Graphic Claims Of Coercion Ahead Of Trial* (2023) *BBC News*. Available at: <https://www.bbc.com/news/world-europe-66581218> (Accessed: 31 August 2023).

³¹ Levine, A. (2022) 'These TikTok Accounts Are Hiding Child Sexual Abuse Material In Plain Sight', *Forbes*, 14 November. Available at: <https://www.forbes.com/sites/alexandralevine/2022/11/11/tiktok-private-csam-child-sexual-abuse-material/?sh=5fa2b8ed3ad9> (Accessed: 31 August 2023).

The Children’s Commissioner (2023)³² published research conducted in the UK. This report draws together research from focus groups with teenagers aged 13-19 and a survey of 1,000 young people aged 16-21. Of the 64% who said that they had ever seen online pornography, the report shows that pornography exposure is widespread and normalised – to the extent children cannot ‘opt-out’ and that the average age at which children first see pornography is 13, but by age nine, 10% had seen pornography, 27% had seen it by age 11. It also shows that young people are frequently exposed to violent pornography, depicting coercive, degrading or pain-inducing sex acts and 79% had encountered violent pornography before the age of 18. Young people expressed concern about the implications of violent pornography on their understanding of the difference between sexual pleasure and harm. Indeed, this report finds that frequent users of pornography are more likely to engage in physically aggressive sex acts. Moreover, pornography is not confined to dedicated adult sites, as it found that Twitter was the online platform where young people were most likely to have seen pornography, followed by Instagram and Snapchat ranking closely after dedicated pornography sites.

A Women’s Aid (2022)³³ study shows that the Irish public is concerned about the pervasiveness and harm of pornography in Irish society, particularly how the exposure to and consumption of pornography is negatively impacting children and young people. The majority believe that it is contributing to gender inequality, sexist double standards, unrealistic sexual expectations, normalisation of requests for sexual images including among children, and directly contributing to coercion and violence against women and girls, including image-based sexual abuse. This study indicates that there is a majority view across all ages that both the government and tech companies need to do more to protect children and young people from exposure to pornography and to do far more, faster, to support victims/survivors of image-based sexual abuse. There is strong support for age-appropriate education for children and young people about sex, relationships, mutuality, consent, and respect as part of school SPHE and RSE curriculum. The research was conducted in October 2022 using the Red Line (a representative sample of the adult population, 18+) and data was weighted across gender, age, region, social class and ethnicity. The National Council for Curriculum and Assessment (NCCA) has finalised an updated SPHE curriculum that is being rolled out for Junior Cycle students from Sept 2023. The new course provides 100 hours of learning (an increase from 70) over the three years of the cycle. The new curriculum will address issues such as consent, the concept of gender identity and the effects of pornography. Schools must be supported to deliver the curriculum in full which would include the appropriate training of teachers in the new course specification. A

³² Children’s Commissioner (2023). ‘A lot of it is actually just abuse’ Young people and pornography. Available at <https://assets.childrenscommissioner.gov.uk/wpuploads/2023/02/cc-a-lot-of-it-is-actually-just-abuse-young-people-and-pornography-updated.pdf>

³³ Women’s Aid (2022). It’s time to talk about porn Irish attitudes on the links between pornography, sexual development, gender inequality and violence against women and girls. Available at https://www.womensaid.ie/app/uploads/2023/05/its_time_to_talk_about_porn_report_womens_aid_november_2022.pdf

consultation is currently underway as part of the process of similarly updating the senior cycle SPHE curriculum.

Wheatley's (2022)³⁴ research conducted in partnership with the NWC, focuses on Social media and online experiences of women in Irish journalism. Drawing on interviews with 36 national-level female journalists, the research explores the emotional and professional burden, as well as the impact on the functioning of democratic governance, the lack of female representation in decision-making roles, and online abuse as an increasing barrier to women's equality in politics. It is particularly shown in the report that journalists could clearly identify the particular topics that would always attract negative online engagement. Some participants singled out Traveller issues and others remarked on migration as a topic that attracted consistently hostile online responses. The report makes recommendations to social media companies to handle content that they deem to be untrue or abusive, to prevent harmful content and to better monitor and verify users. It also provides recommendations for employers to proactively and effectively prepare journalists, handle incidents and ensure there are clear pathways and supports in place. Finally, in terms of policy and legislation, it recommends ensuring meaningful implementation of the objectives and aims of the Online Safety and Media Regulation Act, with a pro-active and sensitive manner to the particular challenges for journalists, and apply pressure on social media platforms to make changes and address the safety of their users and consider penalties.

Ringrose's (2021)³⁵ research on image-based sexual harassment and abuse focusing on young people presents findings from qualitative and quantitative work on digital image-sharing practices with 480 young people aged 12 to 18 years from across the UK. The findings show that non-consensual image-sharing practices were particularly pervasive and consequently normalised and accepted among youth (including unwanted sexual images such as cyberflashing or unsolicited 'dick pics', as well as unwanted solicitation for sexual images such as pressured sexting, and the non-consensual recording, distribution, and/or threat of distribution of sexual images). The aim of the research was to improve the support available for young people by helping parents, teachers, and policymakers to identify and respond to diverse young people's experiences with image-based sexual harassment and abuse. Based on the findings, recommendations are provided for schools, parents and carers, tech companies, and welfare professionals as well as for future research.

Follow up contact can be made to: Ivanna Youtchak ivannay@nwci.ie

³⁴ WHEATLEY (2023). SOCIAL MEDIA AND ONLINE HOSTILITY: EXPERIENCES OF WOMEN IN IRISH JOURNALISM. Available at https://www.nwci.ie/images/uploads/Social_media_and_online_hostility_Experiences_of_women_in_Irish_journalism.pdf

³⁵ Ringrose, J. et al (2021). Understanding and Combatting Youth Experiences of Image-Based Sexual Harassment and Abuse. Available at <https://www.ascl.org.uk/ASCL/media/ASCL/Our%20view/Campaigns/Understanding-and-combatting-youth-experiences-of-image-based-sexual-harassment-and-abuse-full-report.pdf>

Submission from

Athlone Midland Rape Crisis Centre
Dublin Rape Crisis Centre
Galway Rape Crisis Centre
KASA Kilkenny
Sexual Violence Cork
Sligo Rape Crisis Centre
Tullamore Rape Crisis Centre
Wexford Rape Crisis

on developing Ireland's
first binding online safety code
for
video-sharing platform services

September 2023

► Rape Crisis Centres

Rape Crisis Centres (RCCs) provide crisis counselling and long-term therapy to those who have experienced rape, sexual assault and childhood sexual abuse. The services include helplines and associated services, one-to-one counselling, medical, Garda and court accompaniment, education and training programmes, policy interventions, public awareness campaigns to prevent sexual violence and data collection and analysis on trends and issue relating to sexual violence. The work carried out by RCCs has prompted social, political and cultural changes in Ireland.

The following RCCs work together on common issues as members of the Rape Crisis Centre Managers Forum and constitute half of all the Rape Crisis Centres in Ireland.

We are:

1. Athlone Midland Rape Crisis Centre;
2. Dublin Rape Crisis Centre;
3. Galway Rape Crisis Centre;
4. KASA Kilkenny;
5. Sexual Violence Cork;
6. Sligo Rape Crisis Centre;
7. Tipperary Rape Crisis Centre;
8. Tullamore Rape Crisis Centre; and
9. Wexford Rape Crisis.

As frontline service providers, we work with and support people who have been directly affected by sexual violence including online abuse. Through that work, we see the often life-long consequences of the trauma and harm caused by sexual violence of all kinds. We also know from our experience that often this harm is because of digital technology that is used to harass and humiliate.

Eight of our colleague organisations from the Forum; Athlone, Dublin, Galway, Kilkenny, Cork, Sligo, Tullamore and Wexford join with us in making this submission which is informed by the experiences of the women and men accessing these services who are victims of sexual violence.

► About this submission

We welcome the opportunity to contribute to this consultation process. We have structured our responses in the form of answers to the questions set out in the consultation document. In addition, we support the submission being made by the Children's Rights Alliance on behalf of a coalition of organisations including Dublin Rape Crisis Centre. The particular focus of that submission relates to children and young people but is equally applicable to the wider population, in particular those who are additionally vulnerable because of age, gender, relational abuse or other issues.

► Questions and responses

Question 1: What do you think our main priorities and objectives should be in the first binding Online Safety Code for VSPS? What are the main online harms you would like to see it address and why?

Ireland's first binding Online Safety Code (the Code) should be the benchmark for requiring VSPS providers to protect online users from harm by ensuring their services make appropriate use of systems and process to keep users safe. Some of the main priorities and objectives that should be considered are:

- **User Safety and Well-being:**
 - The primary objective should be to safeguard users from various forms of online harm and ensure their safety, well-being and privacy.
- **Platform Responsiveness:**
 - Put time limits in place for providers to remove illegal or harmful content upon identification. Require platforms to impose proportionate sanctions on perpetrators including account suspension and termination.
- **Transparency and Accountability:**
 - Users should know how decisions about content removal are made. The providers should publish regular reports that include content moderation and enforcement actions.
- **User Empowerment:**
 - Promote collaboration between the providers and educational institutions to promote digital literacy. Require providers to promote awareness among users of the avenues of complaint and redress available to them.
- **Regular Review and Update:**
 - Ensure regular reviews and updates to adapt to new challenges in the ever-evolving online environments.

The Code should address a wide range of online harms to create a safer and more secure digital environment. Many of these harms can have a significant negative impact on individuals, communities, and society as a whole. Included in the harms the Code seeks to address should be those outlined in Article 28b of the Audiovisual Media Services Regulation¹ and all the categories of harm set out in the Broadcasting Act 2009 as amended by the Online Safety & Media Regulation Act 2022.²

In particular, the Code should address technology-facilitated gender-based violence (TFGBV). Technology-facilitated GBV refers to any act that is committed, assisted, aggravated or amplified using Information and Communication Technologies (ICTs) or

¹ EU's Audiovisual Media Services Directive: <https://eur-lex.europa.eu/eli/dir/2018/1808/oj>

² Online Safety & Media Regulation 2022: <https://www.irishstatutebook.ie/eli/2022/act/41/enacted/en/print.html>

other digital tools, that results in or is likely to result in physical, sexual, psychological, social, political or economic harm, or other infringements of rights and freedoms.³

In our work, we regularly hear from those using our services that the abuser, who may be known to them or not, posts or threatens to post intimate images of them without their consent to humiliate, intimidate, or blackmail them. Survivors of sexual violence can be subjected to online trolling or negative and abusive commentary which can be incredibly harmful to them personally and can also have a broader effect of deterring victims/survivors from seeking help or reporting their assault. Recently, DRCC launched an anonymous online platform where survivors of sexual violence can share their stories without fear of being trolled. The purpose of We-Speak⁴ is to provide a platform for survivors of sexual violence to reclaim their narrative and safely tell their own stories.

The Platform of Independent Expert Mechanisms on Discrimination and Violence against Women (EDVAW) first thematic paper on the digital dimension of online abuse describes online and technology-facilitated violence against women as having a devastating impact on women and girls, and society generally. It is often experienced as an all-encompassing harm impacting on every aspect of their lives, leading to a form of ‘social rupture’ where lives are divided into ‘before’ and ‘after’ the violence.⁵ These are sentiments echoed by those using RCCs in terms of the devastating impact the online harm had on their lives and why it is so important that they are addressed.

The Code should proceed on the understanding that violence or harm perpetrated online is just as serious as harm perpetrated offline. Perpetrators of technology-facilitated GBV should not be enabled to evade accountability or hide behind a veil of anonymity by reason of weak or inadequate procedures imposed by VSPS providers. Users subjected to technology-facilitated GBV suffer real life impacts and harms and must have clear access to remedies and supports.

The UN Human Rights Council has long-since clarified the principle that human rights protected offline should also be protected online.⁶ Indeed, the UN special Rapporteur on Violence against Women, Its Causes and Consequences warned, in 2018, of the significant risk that the use of ICT without a human rights-based approach and in the absence of the prohibition of online gender-based violence could broaden sexual and gender-based discrimination and violence against women and girls in society even further.⁷

³ See Technology-facilitated Violence against Women: Towards a common definition Report of the meeting of the Expert Group 15-16 November 2022, New York, USA available at <https://www.unwomen.org/sites/default/files/2023-03/Expert-Group-Meeting-report-Technology-facilitated-violence-against-women-en.pdf>

As acknowledged in the expert group report, technology-facilitated violence disproportionately impacts women in all their diversity and gender non-conforming individuals; it is noted that “violence against women” (VAW) can be substituted with “gender-based violence” (GBV), whilst maintaining the common definition describing the phenomenon.

⁴ <https://www.wespeak.ie/>

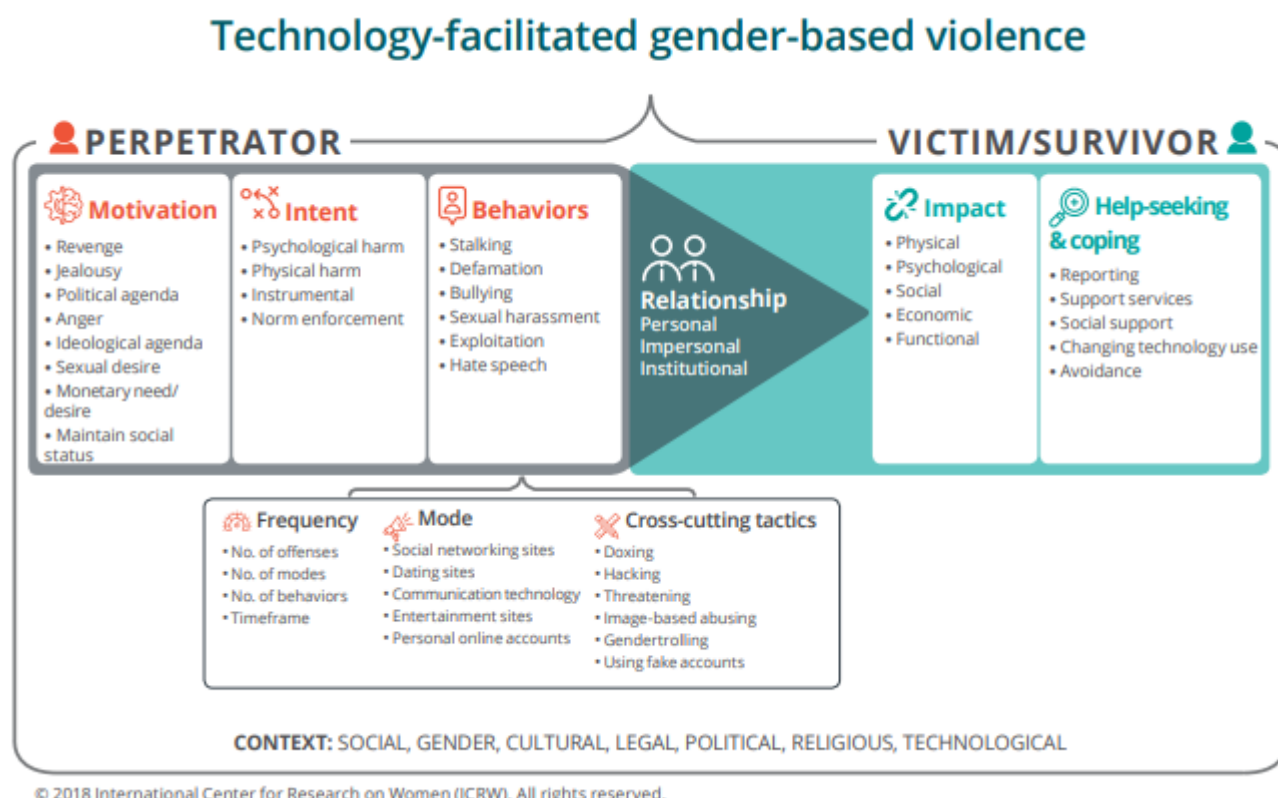
⁵ Platform of Independent Expert Mechanisms on Discrimination and Violence against Women (EDVAW Platform)): *The digital dimension of violence against women as addressed by the seven mechanisms of the EDVAW Platform* (2021), available at: https://www.ohchr.org/sites/default/files/documents/hrbodies/cedaw/statements/2022-12-02/EDVAW-Platform-thematic-paper-on-the-digital-dimension-of-VAW_English.pdf

⁶ Human Rights Council resolution 32/13.

⁷ Report of the Special Rapporteur on Violence against Women, Its Causes and Consequences on online violence against women and girls from a human rights perspective at para 19 available at <https://digitallibrary.un.org/record/1641160?ln=en>

A primary objective of the Code should be to combat and prevent the ever-evolving forms of technology-facilitated GBV, while upholding the right to freedom of expression, including access to information, the right to privacy and data protection, as well as the rights of women that are protected under the international human rights framework.

For the purposes of the Commission’s research and preparation of the draft Code, we refer to the research published by the International Centre for Research on Women who produced the following infographic to help summarise the conceptual framework in which technology-facilitated GBV sits:⁸



Question 2: What types of online harms do you think should attract the most stringent risk mitigation measures by VSPS? How could we evaluate the impact of different types of harms e.g., severity, speed at which harm may be caused? Is there a way of classifying harmful content that you consider it would be useful for us to use?

Using technology and being online is an integral part of everyone’s life, including for education, employment and social interactions. However, if the devices we use become sites of trauma as a result of online abuse, then the knock-on effect of that abuse is one that directly and adversely impacts every aspect of our lives.

⁸ Hinson L, Mueller J, O’Brien-Milne L, Wandera N. (2018). Technology-facilitated gender-based violence: What is it, and how do we measure it? Washington D.C., International Center for Research on Women available at https://www.icrw.org/wp-content/uploads/2018/07/ICRW_TFGBVMarketing_Brief_v8-Web.pdf

Covid-19 exacerbated online and technology-facilitated GBV⁹ and a number of studies suggest that the current prevalence of digital violence is high, mostly impacting young women and girls, women in public life, and people with intersecting identities.¹⁰

¹¹ According to the EU Agency for Fundamental Rights' survey on violence against women (2014), 14% of women in the EU have experienced stalking in the form of offensive or threatening communications since the age of 15.¹² A report commissioned by Women's Aid shows that 1 in 5 young women and 1 in 11 young men in Ireland have suffered intimate relationship abuse. In all cases where women were subjected to intimate relationship abuse, this abuse was perpetrated by a current or former intimate male partner.¹³ According to the report *Toxic Twitter* issued by Amnesty International, 25% of respondents polled across eight countries had received threats, including of sexual violence, physical pain, incitement to suicide and death towards them and their family on Twitter.¹⁴ Plan International, found that more than half of the interviewed 14 000 15 to 25 year old women from 22 different countries said they had been cyberstalked, sent explicit messages and images, or abused online.¹⁵

The harms of key concern to RCC's where the most stringent measures need to be applied relate to:

- Intimate image abuse (IIA)¹⁶
- Technology-facilitated GBV the range and extent of harms which arise online, and which disproportionately impact women, girls and LGBTI individuals.¹⁷
- Cyberbullying and online harassment
- Child Sexual Abuse Material (CSAM), Child Sexual Abuse Imagery (CSAI), child pornography
- Child Sexual Exploitation (CSE) and technology facilitated child sexual exploitation
- Computer generated or drawn content depicting gross child sexual abuse¹⁸
- Non-consensual posting of a person's details on escort websites/OnlyFans etc (whether intimate image abuse involved or not)

⁹ The Ripple Effect: COVID-19 and the Epidemic of Online Abuse by Glitch UK and End Violence Against Women Coalition available at <https://glitchcharity.co.uk/wp-content/uploads/2021/04/Glitch-The-Ripple-Effect-Report-COVID-19-online-abuse.pdf>.

¹⁰ See Practice Brief on Innovation and Prevention of Violence Against Women issued by UN Women, July 2023 available at <https://www.unwomen.org/en/digital-library/publications/2023/07/innovation-and-prevention-of-violence-against-women>

¹¹ The Ripple Effect: COVID-19 and the Epidemic of Online Abuse by Glitch UK and End Violence Against Women Coalition (above). This survey found that "gender was the most often cited reason for online abuse. 48% of respondents reported suffering from gender-based abuse; 21% of respondents reported suffering from abuse related to their gender identity and sexual orientation, followed by 18% for their ethnic background and 10% for their religion and 7% for a disability."

¹² Fundamental Rights Agency (2014), 'Violence against women: an EU-wide survey. Main results report', available at <https://fra.europa.eu/en/publication/2014/violence-againstwomen-eu-wide-survey-main-results-report>

¹³ Women's Aid (2020), 'One in Five Young Women Suffer Intimate Relationship Abuse in Ireland', available at https://www.womensaid.ie/app/uploads/2021/03/one_in_five_women_report_womens_aid_2020.pdf

¹⁴ Amnesty International (2018), 'Toxic Twitter, a toxic place for women', available at www.amnesty.org/en/latest/research/2018/03/online-violence-against-women-chapter-1

¹⁵ Plan International (2020), 'Free to be online? A report on girls' and young women's experiences of online harassment', available at <https://plan-international.org/publications/freetobeonline>

¹⁶ Intimate image abuse has previously been referred to as 'revenge porn' which is widely accepted now as wholly inappropriate.

¹⁷ See Hinson L, Mueller J, O'Brien-Milne L, Wandera N. (2018). Technology-facilitated gender-based violence: What is it, and how do we measure it? Washington D.C., International Center for Research on Women available at https://www.icrw.org/wp-content/uploads/2018/07/ICRW_TFGBVMarketing_Brief_v8-Web.pdf

¹⁸ 9% of CSAM assessed by Hotline.ie in 2021 was computer generated. It would often then contain a disclaimer that 'no child has been harmed in the process'. Hotline.ie 2021 Annual Report, p.15: <https://hotline.ie/library/annual-reports/2022/Hotline.ie-AR21-webready.pdf>

- Doxing - posting a person's private details online such as their address or phone number without their permission and with the aim to cause alarm or distress
- Victims of trafficking being displayed on escort websites
- AI/Computer generated or drawn content depicting a person's identity without consent, especially where it constitutes an intimate image or contains sexual violence.

In 2021, Hotline.ie the national reporting centre for illegal online content received the highest number of reports in one year, 29,794 reports compared with 10,699 the previous year. The images involve a victim(s) who have suffered abuse but who go on to be re-victimised each and every time the image of their abuse is viewed. The 2021 report also included, for the first time, statistics on intimate image abuse (IIA), or the non-consensual sharing of intimate images and videos. Between September 2021 and September 2022, Hotline.ie received 773 reports of suspected IIA. Hotline report that 83% of victims of IIA processed by Hotline.ie are female; 16% male and 1% prefer not to say. DRCC helpline staff have noted an increase in male victims reaching out for assistance and support in recent months.

Evidence also suggests that there is a clear, but often overlooked, overlap between online and offline (or in person / physical) abuse and violence. Perpetrators may target a victim in multiple ways simultaneously. For example, a perpetrator of in person intimate partner abuse may, either during the relationship or at the point their partner ends the relationship, turn to online forms of harassment, abuse and extortion.¹⁹

Platforms and their moderators require specialised training in identifying and understanding domestic, sexual and gender-based violence (DSGBV) and to understand the dynamics of consent, control, coercion and harm. In a survey carried out in the UK regarding online abuse during the Covid-19 pandemic, 83% of respondents who reported one or several incidents of online abuse during COVID-19 felt their complaint(s) had not been properly addressed. This proportion increased to 94% for Black and minoritised women and non-binary people.²⁰

The Code must ensure that illegal material such as child sexual abuse materials and intimate images are removed quickly.

Providers should provide more transparency about their policies related to online abuse, including their position as regards dehumanising language based on gender, ethnicity and other protected categories. Providers should engage actively and regularly with experts in the field of GBV (and child protection) and regularly review and update policies to address new trends, patterns and manifestations of online abuse, including violence against women and people with intersecting identities.

¹⁹ See Hinson L, Mueller J, O'Brien-Milne L, Wandera N. (2018). Technology-facilitated gender-based violence: What is it, and how do we measure it? Washington D.C., International Center for Research on Women at https://www.icrw.org/wp-content/uploads/2018/07/ICRW_TFGBVMarketing_Brief_v8-Web.pdf
See also UN Women Brief: The state of evidence and data collection on technology-facilitated violence against women, 2023 available at <https://www.unwomen.org/en/digital-library/publications/2023/04/brief-the-state-of-evidence-and-data-collection-on-technology-facilitated-violence-against-women> which cites Messing et al.'s interviews with residents of a women's shelter which helped illustrate how technologies are interwoven throughout women's experiences of stalking and abuse, making the distinction between 'offline' and 'online' violence blurry – especially given women's need to continue using digital technologies for their livelihoods and, indeed, to escape situations of violence.

²⁰ The Ripple Effect: COVID-19 and the Epidemic of Online Abuse by Glitch UK and End Violence Against Women Coalition (above).

Some early-stage research is currently investigating the dichotomy between what types of harmful content online platforms seek to curb and what research efforts there are to automatically detect such content. The research paper discusses the mismatch in focus as well as other challenges to be addressed in addressing online harms including fluid policies and platform responsibility and directions for future work.²¹

Question 4: What approach do you think we should take to the level of detail in the Code? What role could non-binding guidance play in supplementing the Code?

We would support option 1 in the consultation paper that the Code should be a very detailed prescriptive Code that could specify in detail the measures the VSPS would be expected to take to address online harms.

Non-binding guidance should not be utilised in a way which would dilute or obfuscate the obligations on providers to remove harmful content within specified, rapid timeframes, to have proportionate and effective age-identification mechanisms in operation, to offer users clear and accessible channels to report harmful content and to take proportionate steps against perpetrators of harm (including suspension and termination).

We believe the Code should address content connected to video content including comments on videos, descriptions of videos or text and images embedded with videos. We would additionally suggest that this extends to links cited in or under video content which leads to harmful content elsewhere on the internet (an obvious example here would be a link to an adult pornography website embedded in content available to children).²²

Question 10: What requirements should the Code include about age verification and age assurance? What sort of content should be shown by default to users who are logged out or in private browsing mode and whose age cannot be verified or assured? What evidence is there about the effectiveness of age estimation techniques? What current practices do you regard as best practice? Where accounts are not age verified should default privacy settings be used, should content default to universal content and should contact by others be more limited?

We support age verification. There must be a level of thoroughness proportionate to the risk of harm due to the nature of the content on the platform. For online users whose age cannot be verified or assured, the default content should prioritise safety and appropriateness and show content that is suitable for all audiences.

²¹ *Detecting Harmful Content on Online Platforms: What Platforms Need vs. Where Research Efforts Go* was accepted to ACM Computing Surveys <https://dl.acm.org/doi/epdf/10.1145/3603399>

²² *Children exposed to 'vast amounts' of inappropriate content online* (September 2022) <https://www.rte.ie/news/2022/0906/1320777-cyber-safety-ireland/>

Question 13: What requirements should the Code contain to ensure that VSPS provide for effective media literacy measures and tools?

VSPS providers should see media literacy measures and tools as a foundational necessity in being involved in the digital world. It is not enough to have good terms and conditions and safety features in place, if the users of the technology are not enabled to actively engage with them. VSPS providers must roll out accessible, age-appropriate educational initiatives to help users understand how to stay safe online, how to respond to online abuse and how to be an active online bystander. Providers should engage the expertise of organisations working in the field of child protection and GBV.

Question 14: How should we ask VSPS providers to address online harms in their terms and conditions in the Code, including the harms addressed under Article 28b? How should key aspects of terms and conditions be brought to users' attention? What examples are there of best practice in relation to terms and conditions including content moderation policies and guidelines?

Transparency and simplicity are key in bringing the terms and conditions to users' attention. The key terms in plain user-friendly language, without the use of jargon or legalese should be prominently displayed during the registration or sign-up process to ensure users see them before proceeding. The use of visual cues like graphics or symbols to draw attention to important aspects or interactive features that require user engagement could be included to ensure the key aspects of the terms and conditions are brought to users' attention.

Strict implementation of a provider's terms and conditions is vital. Rapid Take Down Protocols, together with account suspensions and terminations, will send an unambiguous message to perpetrators and potential perpetrators that such abuse will not be tolerated, which in itself can have a preventative impact.

The Code should also require providers to actively establish and utilise systems to identify repeat offenders of online abuse. Anonymity is a key tool utilized by persons intent on causing harm online. Providers should be required to take steps to make it far more difficult for accounts that have been the subject of a ban or termination to resurface as a new account.

Question 16: What requirements should the Code include about procedures for complaint-handling and resolution, including out-of-court redress or alternative-dispute resolution processes? To what extent should these requirements align with similar requirements in the DSA? What current practices could be regarded as best practice? How frequently should VSPS providers be obliged to report to the Commission on their complaint handling systems and what should those reports contain? Should there be a maximum time-period for VSPS providers to handle user complaints and if so, what should that period be?

Time is of the essence for intimate image abuse (IIA), the longer it takes to remove content, the greater the risk of repeat victimisation. Once an intimate image is online, it is very easy to copy, save, replicate and spread. Action needs to be immediate. The user should be enabled to make their complaint directly to the VSPS provider, who should proceed based

on accepting the truth of a statement of non-consent and should promptly remove the content. **A precautionary approach in favour of removal is appropriate here.**

Whether consent was forthcoming or not at the time the image was uploaded is irrelevant to the question of removal as consent can be revoked. The key facts relevant to the platforms should be whether the image in question is of the complainant. The facts and evidence around consent (if non-consent is contested by the user who posted the image/video) are primarily relevant to any criminal investigation An Garda Síochána undertake, and platforms should preserve all relevant evidence for same. Such questions may also be relevant to any decision the provider takes as regards a sanction against the user who posted/hosted the disputed image or in respect of any review that user may take against a decision to take down or to suspend/terminate their account. However, as time is so vitally of the essence in the case of IIA, removal on a precautionary (and possibly temporary) basis should be the default with providers conducting any more detailed factual investigations only thereafter.

If the offending content is not taken down or a notice not complied with in a take-down timeline specified in the Code, then the user should have access to an accessible and effective complaint mechanism. The complaint mechanism should offer a very prompt internal review of the initial decision so that legitimate requests to takedown harmful content are not unduly delayed which would in turn result in serious and escalating harm to the user/victim. The user should also be offered an avenue to seek an external review of a complaint to an independent body such as the Commission.

In the case of harmful content of a sexual or intimate nature such as IIA²³ and CSAM, the time-frames in question for both an initial moderation decision on a take-down request and a complaint/request for a review should be in the order of hours not days i.e. an initial decision should be taken within 12-24 hours and a review decision should be the same in cases whether the disputed content remains online. Once, and for so long as, the disputed content is removed (even temporarily), longer timeframes may be acceptable for the processing of final decisions and reviews/complaints.

The Code should not require any person to engage in mediation with a perpetrator of harm or GBV. Out of court redress or alternative dispute resolution processes such as mediation may be relevant and appropriate to a dispute between users and VSPS providers but only in cases where the user consents to such processes. Education and awareness raising of user's rights in this respect should be rolled-out.

VSPS providers should submit quarterly reports on the measures and actions they have put in place to combat harmful and inappropriate content. These reports should contain comprehensive data and details as regards users experience of the VSPS provider's platform and complaint handling systems. This data should include details of the number of take-down requests received, the number acted on and the number dismissed, the number of account suspensions and terminations, the number of user complaints received and the outcome of same. All data should be anonymised and disaggregated by age and gender of perpetrator(s) and victim(s) where known and the nature of the disputed content. In this ever evolving and growing digital space, such data and detail is necessary

²³ Content which includes a person's image and contact details or a suggestion that the person is seeking or available for intimate contact where that person has not consented should also be treated with the utmost urgency in moderation and take-down decisions.

to enable the Commission, researchers and users to understand what is working, what is not working and what changes and updates are necessary.

VSPS providers should also be required to report on their digital literacy efforts and training initiatives, to include details of the nature of specialised training moderators and staff involved in design and safety features receive in relation to child protection and GBV matters. Ultimately the reporting and resolution mechanisms must be effective, transparent, easy to access and easy to use.

Question 17: What approach do you think the Code should take to ensuring that the safety measures we ask VSPS providers to take are accessible to people with disabilities?

Ensuring that the safety measures required by the Code are accessible to people with disabilities is essential to creating an inclusive online environment. VSPS providers should be required to adhere to recognised accessibility standards such as the Web Content Accessibility Guidelines (WCAG) 2.1,²⁴ alt text for images, keyboard navigation, and screen reader compatibility to ensure that their safety measures are accessible. Terms and conditions and reporting procedures should be available in alternative formats i.e., audio, braille, or plain text for users with various disabilities.

Question 18: What approach do you think the Code should take to risk assessments and safety by design? Are there any examples you can point us towards which you consider to be best practice?

Security-by-design, privacy-by-design and user safety considerations should be standard requirements in product/service development by VSPS. Impact/risk assessment frameworks should be applied with appropriate checks and balances.

Question 20: What approach do you think we should take in the Code to address feeds which cause harm because of the aggregate impact of the content they provide access to? Are there current practices which you consider to be best practice in this regard?

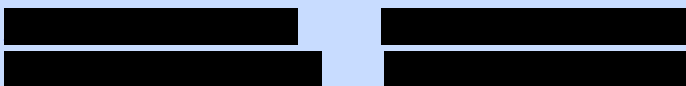
It is vital for providers (and their design staff and moderators) to have an evidentially informed understanding of GBV to be able to design safety features and assess user complaints effectively. For example, most of the literature tends to discuss incidents of abuse as single events requiring a moderator to consider a single video/comment/post and to make a decision on that individual item. However, this ignores the multiple acts of abuse and violence a person may experience online, potentially across platforms, and simultaneously offline. A perpetrator of abuse may engage in stalking, defamation, bullying, sexual harassment, exploitation and/or hate speech. These may be repeated behaviours carried out across multiple platforms. Alongside speedy moderation processes for single events which obviously constitute abuse or illegal behaviour, there should also be channels for users to report accounts carrying out multiple behaviours which culminate in abuse against a user (or group of users) even where each single incident in isolation may not establish the abuse.

²⁴ Web Content Accessibility Guidelines (WCAG) 2.1: <https://www.w3.org/TR/WCAG21/>

Question 22: What compliance monitoring and reporting arrangements should we include in the Code?

To be effective, all VSPS providers need to be subject to the Online Safety Code – any platform that seeks to evade the obligations under the Code undermines the objective of making online activity safer for children and all users. Those intent on perpetrating harm will favour the platforms that seek to remain outside the Code or other Regulations (perpetrators are known to employ ‘*platform hopping tactics*’²⁵). RCCs are concerned that the effective progression and implementation of these and similar measures may be greatly undermined should platforms bring about delays through avoidable court proceedings on technical points or matters that could be pre-empted and avoided at this early juncture.²⁶ This consultation is key to seeking to address all concerns and viewpoints of all relevant actors in the hope of avoiding such delays. Equally, RCCs assumes platforms will adopt a reasonable and proactive approach to assisting the Commission adopt an online safety code that is effective, practicable and acceptable to all.

For further information, on behalf of the group, please contact:



National 24-Hour Helpline: 1800 77 88 88

²⁵ See Hotline.ie Annual Report 2021 p.16. Hotline.ie also note perpetrators use ‘*breadcrumbing*’ tactics to essentially leave an innocuous trail of clues across various websites to eventually lead to CSAM i.e. as a form of distribution.

²⁶ See EU safety laws start to bite for TikTok, Instagram and others, BBC News 25 August 2023 ([here](#)) which details: “Retailers Zalando and Amazon have mounted legal action to contest their designation as a very large online platform. Amazon argues they are not the largest retailer in any of the EU countries where they operate. Nevertheless, Amazon has taken steps to comply with the act and has “created a new channel for submitting notices against suspected illegal products and content”. Zalando told the BBC it will be compliant with the act.”

Subject:

FW: Private and Confidential - Online Safety Code Submission

From:

Sent: Monday, September 4, 2023 1:35 PM

To: VSPS Regulation <vspsregulation@cnam.ie>

Subject: Private and Confidential - Online Safety Code Submission

You don't often get email from. [Learn why this is important](#)

To Whom it may concern,

I have worked in the area of online safety for a number of years and from these experiences I propose the following:

1) The Irish domain registry (formerly IEDR) now known as [weare.ie](#) cannot be trusted with regulation of dot ie domain names. I have yet to see them take action against any hate sites that I have reported. They do not enforce their own rules on illegal activity such as hate speech, fraud or disinformation. The company either needs to be sanctioned or brought into public ownership.

2) Some public bodies and semi-state agencies do not observe [section 42 of the 2014 Irish Human Rights and Equality Act](#) with regard to their intellectual property. This can entail their intellectual property being used to spread hate online such as YouTube videos, photos, logos, uniforms, trademarks etc. Both the Houses of the Oireachtas and the Dept of Defence have been the subject of complaints in this regard and have done next to nothing. Social media companies that publish trademarked and copyrighted material that spreads hate should be served with a High Court injunction if needs be. The costs involved will soon dissuade the social media companies from refusing to comply as they currently do. The IHREC and its CEO have been aware of this issue and apparently have done nothing.

<https://www.ipoi.gov.ie/en/understanding-ip/ip-infringement/enforcing-your-ip/>

3) Social media companies have been compared to the tobacco industry of the 1960s that downplayed the serious harms of smoking. Social media companies and websites that fail to comply with regulation should be geo-blocked similar to the way RT (Russia Today) has been blocked in the EU since Russia's full-scale invasion of Ukraine in 2022.

4) There needs to be a campaign of educating the public in how to behave responsibly online e.g. not to engage with bad actors or spread disinformation and hate. Gardaí too are sorely lacking in the policing of online activity. Fraud, hate speech and threats are regularly ignored by An Garda Síochána in my experience. [Hotline.ie](#) is in my experience a complete work of fiction in these areas. Industry cannot be trusted to police itself. The [Garda Hate Crime Reporting Tool](#) has been a complete disappointment. Videos and social media posts I reported last November are still online. Gardaí have shown little enthusiasm in preventing online crime such as fraud. I was forced to contact a hosting company myself in the Netherlands myself in order for a number of websites set up by an Irish citizen to be taken offline.

5) Social media posts and website content that can be published online in a few seconds should also be able to be removed in a number of seconds. Human eyeballs will need to verify content removal, not AI. Gardaí have spoken at Oireachtas committees in the past [about removing harmful content swiftly](#) but have failed to follow up themselves in this regard. If I'm not mistaken, AGS did not mention content removal in its [submission to the dept](#) in the drafting of the Online Safety Bill.

I do not consent for my name to be made public so I'll just use my initials.

Regards,

SM

CONFIDENTIALITY NOTICE

This email message, including all the attachments, is for the sole use of the intended recipient(s) and contains confidential and privileged information. Unauthorised use or disclosure is prohibited. If you are not the intended recipient, you may not use, disclose, copy or disseminate this information; and please contact the sender immediately by reply email and destroy all copies of the original message, including attachments.



Samaritans Ireland submission to Coimisiún na Meán's Call for Inputs: Online Safety

Developing Ireland's First Binding Online Safety Code for Video-Sharing Platform Services

Samaritans Ireland welcomes the opportunity to respond to Coimisiún na Meán's call for input on the development of *Ireland's First Binding Online Safety Code for Video-Sharing Platform Services*.

Samaritans Ireland is the only all-island 24-hour emotional support helpline. Through over 2,000 listening volunteers, we respond to around 1,500 calls for help every day. We believe in the power of compassionate and non-judgemental listening to give people a safe place to work through their problems.

In 2019, following the death of Molly Russell¹, our Samaritans Central Charity (SCC) colleagues in Great Britain, in collaboration with the UK government and some of the largest tech platforms, established an Online Excellence Programme with the aim of promoting good practice around self-harm and suicide content online. This includes an advisory service for professionals and platforms dealing with self-harm and suicide content online, a published best-practice guidance document² for platforms hosting user-generated self-harm and suicide content, a programme of research to better understand the risks and benefits for users accessing this material, and online user resources to support individuals to talk about suicide and self-harm safely online.

Samaritans Ireland welcomes the opportunity to partake in this consultation. We have focused on questions most relevant to our work and area of expertise. As part of our own response process, we created an informal alliance with other expert voices in the field – *Headline, National Office for Suicide Prevention (NOSP), National Suicide Research Foundation (NSRF)*, and the *Department of Health's Mental Health Unit*. While our response and theirs are all individual, each submission has full support from our alliance.

We appreciate a process to determine these codes relating to suicide and self-harm may be difficult or triggering for any Coimisiún na Meán staff involved and suggest referring to our resource on supporting the wellbeing of staff working with self-harm and suicide content which gives advice on managing wellbeing when viewing potentially distressing content as appropriate.³

¹ <https://www.bbc.com/news/av/uk-50186418>

² Samaritans (2020) Samaritans' industry guidelines: Guidelines for sites and platforms hosting user-generated content

³ Samaritans (2020) Samaritans' guidelines for supporting the wellbeing of staff working with self-harm and suicide content

Question 1: What do you think our main priorities and objectives should be in the first binding Online Safety Code for VSPS? What are the main online harms* you would like to see it address and why?

As a suicide prevention charity, Samaritans Ireland prioritises online harms relating to mental health and specifically self-harm and suicide. Equally, Samaritans Ireland recognises the immensity of areas the Online Safety Codes will be required to cover and has identified three main objectives the codes should address that could apply across all online harms.

- I. Ensure VSPS are minimising harmful content while maximising opportunities for help and support
- II. Hold platforms accountable to ensure their moderators are appropriately skilled and supported to ensure moderation takes place in a manner which does not further stigmatise vulnerable people, like those with self-harm or suicidal thoughts and/or experiences
- III. Outline compliance monitoring and reporting requirements of content moderators as a key way to monitor overall internet safety. The monitoring/report should include specific measures for platforms to ensure the good mental health and wellbeing of people who review/moderate potentially harmful content to ensure they are able to operate at full capacity and effectively remove/reduce harmful or potentially harmful online content.

Samaritans Ireland knows the internet has the potential to be a powerful tool for suicide prevention. It can provide a space of belonging by offering an opportunity to connect with other people who have similar experiences.⁴ It can also provide access to content that can be distressing, triggering⁵ and instructive⁶ and may act to encourage, maintain or exacerbate self-harm and suicidal behaviours.⁷ Other risks include contagion effects caused by over identification with the user who posts the content and imitative and 'copycat' suicides when detailed information about methods is presented.⁸

* [Note from Coimisiún na Meán] Please remember that when we refer to 'online harms' and 'online harm' in this document this includes harm that can be caused by harmful online content, illegal content, inappropriate content, and commercial communications collectively.

⁴ Lavis, A., & Winter, R. (2020). #Online harms or benefits? An ethnographic analysis of the positives and negatives of peer-support around self-harm on social media. *Journal Of Child Psychology And Psychiatry*, 61(8), 842-854.

⁵ Arendt, F., Scherr, S., & Romer, D. (2019). Effects of exposure to self-harm on social media: Evidence from a two-wave panel study among young adults. *New Media and Society*, 21, 2422–2442.

⁶ Biddle, Lucy, Jane Derges, Becky Mars, Jon Heron, Jenny L. Donovan, John Potokar, Martyn Piper, Clare Wyllie, and David Gunnell. "Suicide and the Internet: Changes in the accessibility of suicide-related information between 2007 and 2014." *Journal of Affective Disorders* 190 (2016): 370-375.

⁷ Arendt et al., 2019; Marchant et al., 2017; Niedzwiedz et al., 2014; Biddle et al., 2012; Lavis & Winter, 2020

⁸ Niederkrotenthaler et al, Association between suicide reporting in the media and suicide: systematic review and meta-analysis, 2020

While suicide and self-harm are complex and rarely caused by one thing, in many cases the internet is involved: a 2017 inquiry into suicides of young people in the UK found suicide-related internet use in nearly 26% of deaths in under 20s, and 13% of deaths in 20-24 year olds.⁹

Samaritans Ireland is concerned that harmful content relating to self-harm and suicide is too accessible online and would like to see this content minimised while opportunities for support and help online are maximised. The importance of the internet's role in suicide prevention means it is vital for online platforms to take responsibility for both protecting positive supporting spaces and preventing harmful content. Samaritans Ireland believes that it is critical to protect people of all ages from suicide and self-harm content that's legal but extremely harmful – on both large and small online platforms, specifically those with user-generated content (UGC).

Question 2: What types of online harms do you think should attract the most stringent risk mitigation measures by VSPS? How could we evaluate the impact of different types of harms e.g. severity, speed at which harm may be caused? Is there a way of classifying harmful content that you consider it would be useful for us to use?

All online harms relating to self-harm or suicide should be managed with the highest priority regardless of the intent of the post. As time spent on the internet has increased, so has the speed in which trends and information can travel. In recent years self-harm and/or suicide 'games' or 'challenges' have emerged. Research indicates that novel online risks to mental health, such as pro-suicide games or messages, can circulate quickly and globally¹⁰ and that social media 'games' like the Blue Whale challenge¹¹, can glorify self-harm and suicide, and amplify suicide contagion among vulnerable cohorts¹².

Some suicide and self-harm content is in the 'grey' area and is not easily defined. Communicating online about feelings of suicide or self-harm can be part of a person's recovery¹³, offering support¹⁴ and allowing feelings to be shared without judgement and it is important that supportive content in these spaces is not inadvertently removed.

⁹ Appleby, L. et al., (2017). Suicide by Children and Young People. National Confidential Inquiry into Suicide and Homicide by People with Mental Illness (NCISH). (Manchester: University of Manchester, 2017).

¹⁰ Sumner SA, Ferguson B, Bason B, Dink J, Yard E, Hertz M, Hilkert B, Holland K, Mercado-Crespo M, Tang S, Jones CM. Association of Online Risk Factors With Subsequent Youth Suicide-Related Behaviors in the US. *JAMA Netw Open*. 2021 Sep 1;4(9):e2125860

¹¹ Khasawneh, A., Madathil, K.C., Dixon, E., Wiśniewski, P., Zinzow, H. and Roth, R., 2020. Examining the Self-Harm and Suicide Contagion Effects of the Blue Whale Challenge on YouTube and Twitter: Qualitative Study. *JMIR Mental Health*, 7(6), p.e15973.

¹² Upadhyaya M, Kozman M. The Blue Whale Challenge, social media, self-harm, and suicide contagion. *Prim Care Companion CNS Disord*. 2022;24(5):22cr03314

¹³ Brown, R. C., et al. (2020). "I just finally wanted to belong somewhere" — Qualitative Analysis of Experiences with Posting Pictures of Self-Injury on Instagram', *Front Psychiatry*, 11.

¹⁴ Lavis, A. and Winter, R., 2020. # Online harms or benefits? An ethnographic analysis of the positives and negatives of peer-support around self-harm on social media. *Journal of child psychology and psychiatry*.

People experiencing suicidal feelings or struggling with self-harm are likely to be more vulnerable and at greater risk of harm from legal but harmful suicide and self-harm content.

However, suicidal ideation can quickly fluctuate, sometimes over the course of a single day,¹⁵ meaning that it is difficult to identify who is more vulnerable to suicide and self-harm content: for this reason legal but harmful suicide and self-harm content needs to be regulated across all platforms for people of all ages.

SCC's research with the University of Bristol found that people who use the internet to find out about suicide are likely to be vulnerable and in need of support at that point. Those experiencing high levels of distress show purposeful browsing, looking specifically for information on methods of harm¹⁶. Suicidal people using the internet for suicide-related purposes experience higher levels of suicidality and depression than suicidal people who did not use the internet for this purpose¹⁷.

The Online Safety Codes are an opportunity to ensure these vulnerable users' access to harmful content is minimised, whilst still accessing supportive online spaces with high-quality signposting to support.

SCC recently conducted research (not yet published) with people with lived experience of self-harm and suicide which gave some insight into the impact harmful online self-harm and suicide content can have on adults:

"People need a place where they can express how they feel without backlash. [A place where] I can share how I feel while getting support from peers and not being bombarded with triggering images of open wounds when I am at my most vulnerable" (Adult aged 25-34).¹⁸

Samaritans Ireland has gained valuable insights from our Online Excellence Programme in the UK as to how platform design, systems and processes can be shaped to enhance the safety of their users, and have developed guidelines for the tech industry in managing user-generated suicide and self-harm content, in conjunction with academics, experts and individuals with lived experience.¹⁹ This includes processes for removing detailed information on suicide/self-harm methods; turning off algorithms that push harmful content related to suicide/self-harm; using age and sensitivity warnings; prioritising and promoting positive and helpful content; and effective moderation processes.

Samaritans Ireland would be interested in speaking further with Coimisiún na Meán on establishing thresholds for harm, which should be co-designed with social media users of all ages as well as mental health and industry experts.

¹⁵ Kleiman, Evan M., Brianna J. Turner, Szymon Fedor, Eleanor E. Beale, Jeff C. Huffman, and Matthew K. Nock. "Examination of real-time fluctuations in suicidal ideation and its risk factors: Results from two ecological momentary assessment studies." *Journal of abnormal psychology* 126, no. 6 (2017): 726.

¹⁶ Biddle et al, Suicide and Self-Harm Related Internet Use: A Cross-Sectional Study and Clinician Focus Groups, 2017

¹⁷ Niederkrotenthaler, Thomas, Anna Haider, Benedikt Till, Katherine Mok, and Jane Pirkis. Comparison of suicidal people who use the internet for suicide-related reasons and those who do not. *Crisis* (2016)

¹⁸ Samaritans (2021) Unpublished research. Further information available on request.

¹⁹ Samaritans industry guidelines for managing self-harm and suicide content <https://www.samaritans.org/about-samaritans/research-policy/internet-suicide/guidelines-tech-industry/>

Question 3: Do you have reports, academic studies or other relevant independent research that would support your views? If you do, please share them with us with links to relevant reports, studies, or research.

- Samaritans' Online Safety Industry Guidelines for managing self-harm and suicide content - https://media.samaritans.org/documents/Online_Safety_Guidance.pdf
- Samaritans' co-designed online safety resources for staying safe online when finding support for themselves, or when trying to support others- <https://www.samaritans.org/ireland/about-samaritans/research-policy/internet-suicide/online-safety-resources/>
- Samaritans' guidance for practitioners for how to talk to people who could be at risk of suicide or self-harm about their online activity - <https://www.samaritans.org/ireland/about-samaritans/research-policy/internet-suicide/internet-safety-practitioners/>
- Samaritans' guidance for parents, carers and family members about how to talk to their children about self-harm and suicide content online - <https://www.samaritans.org/ireland/about-samaritans/research-policy/internet-suicide/talking-to-your-child-about-self-harm-and-suicide-content-online/>
- Samaritans' guidance for implementing effective content moderation for self-harm and suicide - <https://www.samaritans.org/ireland/about-samaritans/research-policy/internet-suicide/guidelines-tech-industry/effective-content-moderation/>
- Samaritans' and Swansea University research (2022): How social media users experience self-harm and suicide content - https://media.samaritans.org/documents/Samaritans_How_social_media_users_experience_self-harm_and_suicide_content_WEB_v3.pdf
- The Harmful Impact of Online Content - a Literature Review (November 2020) from the National Suicide Research Foundation - <https://www.hse.ie/eng/services/list/4/mental-health-services/connecting-for-life/publications/the-harmful-impact-of-online-content-a-literature-review.html>
- NSRF Updated Literature Review (August 2023): The Harmful Impact of Suicide and Self-Harm Content Online: A Review of the Literature (NOT YET PUBLISHED PUBLICALLY)

Question 4: What approach do you think we should take to the level of detail in the Code? What role could non-binding guidance play in supplementing the Code?

It is imperative the codes be robust to ensure harmful content is minimised while positive and helpful content is maximised. An article in the Lancet indicated, “Protection and safety frameworks, in addition to voluntary industry codes of conduct to prevent normalisation of harmful behaviour related to suicide and self-harm should be considered.”²⁰ At minimum, Samaritans Ireland would like to see all platforms adopt co-designed guidance, like Samaritans’ Online Safety Industry Guidelines for managing self-harm and suicide content -

https://media.samaritans.org/documents/Online_Safety_Guidance.pdf

Question 5: What do you think would be the most effective structure for the Code? What are the most important factors we should consider when we decide how to structure the Code?

A key aspect of suicide prevention is the reduction of access to means and reducing the availability of harmful and instructive information is one way of achieving this. No caveats around tackling harmful suicide and self-harm content (size of platform, age of user) should be established that will diminish the code’s ability to tackle harmful content in this space. Samaritans Ireland is especially concerned that users of all ages, not just minors, are protected by the codes. A Samaritans’ supporter told us, *“Anyone and everyone who is at risk of even considering suicide needs the online help to prevent them finding the information or impetus they may be looking for to take their own life. I know that every attempt my brother considered at ending his life - from his early 20s to when he died in April aged 40 - was based on extensive online research. It was all too easy for him to find step by step instructions so he could evaluate the effectiveness and potential impact of various approaches, and most recently - given he had no medical background - it was purely his ability to work out the quantities of various drugs, and likely impact of taking them in combination, that equipped him to end his life.”*

Furthermore, it is accepted within the Bill that repeated viewing of harmful content carries its own risk of harm including the being desensitised and/or normalising harmful content. This risk is particularly relevant to the persons who designated service providers are relying on to review content to support moderation and adherence to online safety codes. Samaritans Ireland is concerned both for the mental wellbeing of these individuals as well as for the wider implications to the public should their ability to fulfil their duties be diminished due to desensitisation.

As such, Samaritans has developed best practice guidelines for staff of online platforms, based around supports required and good work practices. We believe inclusion of these in the Online Safety Codes will improve overall safety online and in particular will help a vulnerable and critical group of workers.

²⁰ The Lancet. Social media, screen time, and young people's mental health. *Lancet*. 2019;393(10172):611

Question 9: How should we ask VSPS providers to introduce and design a flagging mechanism in the Code? How can we ensure that VSPS providers introduce the mechanism in a user-friendly and transparent way? How should we ask VSP Providers to report the decisions they've made on content after it has been flagged? To what extent should we align the Code with similar provisions on flagging in the DSA?

It is important that users are equipped with the skills they need to stay safe online. However, all VSPS providers must take responsibility for ensuring the safety of their users, taking appropriate action on self-harm and suicide content that could be harmful.

From Samaritans' user research and SCC's advisory service engagement, Samaritans Ireland is aware that users reporting content often receive poor responses with limited support provided and little or slow action to remove or address the reported content.

A 2023 empirical investigation determined the half-life (or lifespan) of social media posts on different platforms: Snapchat (0min), Twitter (24 min), Facebook (105 min), Instagram (20 h), LinkedIn (24 h), YouTube (8.8 d), and Pinterest (3.75 mo).²¹ A lower half-life means that most harm happens right after the content is posted, and content moderation needs to be performed quickly to be effective. A recent report examining the likely effectiveness of the DSA with regards to regulating highly viral online content found the key to moderation success seem to be appointing trusted flaggers, developing an effective tool for reporting harmful content across platforms, and correctly timing the reaction time for moderation.²²

Some suicide and self-harm content is in the 'grey' area and is not easily defined. Ultimately, while speed of removal is important, any technological interventions to tackle harmful suicide and self-harm content must be underpinned by effective and nuanced human moderation.

Online safety codes should seek to empower these moderators to contribute to a safer, less harmful environment by acting on both content but also algorithms which generate user issues. The assessment of Complaints Handling should be swift and include transparency on the algorithms used in presenting the flagged content and any patterns in these complaints themselves.

²¹ Graffius, Scott. (2023). Lifespan (Half-Life) of Social Media Posts: Update for 2023. 10.13140/RG.2.2.19783.98722.

²² Schneider, Philipp J., and Marian-Andrei RizoIU. "The Effectiveness of Moderating Harmful Online Content." *Proceedings of the National Academy of Sciences* 120, no. 24 (2023).

Question 10: What requirements should the Code include about age verification and age assurance? What sort of content should be shown by default to users who are logged out or in private browsing mode and whose age cannot be verified or assured? What evidence is there about the effectiveness of age estimation techniques? What current practices do you regard as best practice? Where accounts are not age verified should default privacy settings be used, should content default to universal content and should contact by others be more limited?

Samaritans Ireland has gained valuable insights from our Online Excellence Programme in the UK as to how platform design, systems and processes can be shaped to enhance the safety of their users, including using age and sensitivity warnings, prioritising and promoting positive and helpful content, and effective moderation processes.

In Samaritans Ireland's view the Code should have a duty of care to all internet users, regardless of their age and believe all VSPS, regardless of reach and functionality, should be required to remove suicide and self-harm content that is harmful to children and adults.

Whilst it is imperative that children are kept safe online, suicide and self-harm content affects people of all ages. A UK study that looked at deaths by suicide between 2011-2015, found 151 patients who died by suicide were known to have visited websites that encouraged suicide or shared information about methods of harm. 124 were aged over 25.²³ This data was based on clinical reports and is likely to underestimate the true extent to which the internet plays a role in suicides.

In a population survey of 21 year olds, conducted by Samaritans Central Charity and the University of Bristol, almost 75% of the participants who had attempted suicide reported using the internet for a suicide-related purpose; whilst most were seeking help and support, one in five had accessed sites that provided information about methods of harm.²⁴

Additionally, research looking at online support groups found associations between the use of these spaces and suicidal feelings are not limited to younger users, but are also present for people aged 30 to 59.²⁵ It is also important to consider that media literacy is increasing and digital natives are aging with the internet, meaning more 'older' people will find their way online in greater numbers in the near future.

"Harmful and accessible suicide and self-harm online content can be harmful at any age. I am in my fifties and would be tempted to act on this information if I felt suicidal again." – Samaritans' supporter

²³ The National Confidential Inquiry into Suicide and Homicide by People with Mental Illness (NCISH) (2017)

²⁴ Biddle, L., Derges, J., Gunnell, D., Stace, S., Morrissey, J. (2016). Priorities for suicide prevention: balancing the risks and opportunities of internet use. University of Bristol/Samaritans

²⁵ Scherr, Reinemann. First do no harm: Cross-sectional and longitudinal evidence for the impact of individual suicidality on the use of online health forums and support groups (2016)

Question 13: What requirements should the Code contain to ensure that VSPS provide for effective media literacy measures and tools?

User education and media literacy is a key facet of online safety and Samaritans Ireland would point to the current Online Safety Bill currently in Westminster whereby media literacy is underpinned by “an awareness of the impact material may have”²⁶ – this is a key principle of speaking safely about suicide and self-harm online. Samaritans have co-produced a range of user resources with young people with lived experience and would also welcome the opportunity to engage further in this area.²⁷ Extensive engagement with other relevant stakeholders like Media Literacy Ireland and National Adult Literacy Agency will be important to determine specific requirements for VSPS providers.

Question 14: How should we ask VSPS providers to address online harms in their terms and conditions in the Code, including the harms addressed under Article 28b? How should key aspects of terms and conditions be brought to users’ attention? What examples are there of best practice in relation to terms and conditions including content moderation policies and guidelines?

Unlike other areas of online harms, content relating to self-harm and suicide often requires a more nuanced approach. Our research shows that some content on issues like self-harm and suicide can be easily identified as dangerous while other content is recognised to be an important source of support for individuals. When appropriately and ethically regulated, the online environment can provide a supportive forum for people to seek help when they have suicidal thoughts, as well as to interact and build relationships that could help build their emotional resilience.

As laid out in our guidelines – all sites and platforms should take proactive steps to understand the potential benefits and risks associated with self-harm and suicide content online and how it applies to their site while also acknowledging the impact of self-harm and suicide content is complex.

Whilst there are some types of content that are obviously harmful, other types require more nuanced thinking and judgement on what is appropriate for the platform. What can be helpful for one user can be triggering to others. A user’s experience of how harmful content is may also depend on factors such as their current level of distress and the volume of self-harm and suicide content they view. Understanding the potential risks and benefits to users is critical. We believe the definition of ‘online harms’ should allow for these nuances in both the creation/uploading as well as sharing and consumption of the content.

²⁶ Online Safety Bill. Westminster. <https://bills.parliament.uk/bills/3137>

²⁷ <https://www.samaritans.org/about-samaritans/research-policy/internet-suicide/online-safety-resources/>

Samaritans Ireland acknowledges this is a risk and would encourage the development/adoption of guidelines and policies to encourage safe posting and also ensure extensive training is provided to all moderators so any necessary removal of content is conducted in ways which will not further stigmatise those with self-harm or suicidal thoughts and/or experiences. The efficacy of guidelines has already been seen through the widespread adoption of Samaritans' Media Guidelines²⁸ by journalists and organisations in how they report on suicide leading to a reduction in sensationalised or dangerous news stories.

Question 15: How should we ask VSPS providers to address content moderation in the Code? Are there any current practices which you consider to be best practice? How should we address automated content detection and moderation in the Code?

It is important the public be made aware of how to safely talk and post about sensitive topics online as the internet can be a key place individuals seek help and share their own mental health stories. Some suicide and self-harm content is in the 'grey' area and is not easily defined. Samaritans Ireland would welcome the opportunity to speak further with Coimisiún na Meán on how to ensure that supportive suicide and self-harm content is not inadvertently removed. However, ultimately any technological interventions to tackle harmful suicide and self-harm content must be underpinned by effective and nuanced human moderation.

Platforms should be held accountable to ensure their moderators are appropriately skilled and supported to ensure moderation takes place in a manner which does not further stigmatise those with self-harm or suicidal thoughts and/or experiences. Furthermore, in assessing compliance with online safety codes, it is important registered services be mindful of the impact some self-harm and suicide content may have on staff who are tasked with moderating the platforms and the supports available for these staff are rigorously assessed.

Samaritans has provided 24/7 support for over 50 years to callers across Ireland, many of whom were/are suicidal. But equally as important, we have continuously supported the tens of thousands of volunteers who take those calls. While maintaining caller confidentiality, Samaritans has developed substantial debriefing policies to ensure our listeners are able to receive support themselves after a difficult call or shift. The challenges of equipping volunteers with appropriate training and supported to be able to handle the most distressing of calls is an area that Samaritans has always invested in.

²⁸ <https://www.samaritans.org/ireland/samaritans-ireland/about/ireland-media-guidelines-and-online-safety/media-guidelines-in-ireland/>

Samaritans Ireland feel the registered service providers must have a duty to support all staff who undertake moderation of harmful content. Online safety codes should seek to empower these staff to contribute to a safer, less harmful environment by acting on both content but also algorithms which generate user issues. The assessment of Complaints Handling should include transparency on the algorithms used in presenting the flagged content and any patterns in these complaints themselves.

Question 18: What approach do you think the Code should take to risk assessments and safety by design? Are there any examples you can point us towards which you consider to be best practice?

Samaritans Ireland has gained valuable insights from our Online Excellence Programme in the UK as to how platform design, systems and processes can be shaped to enhance the safety of their users, and have developed guidelines for the tech industry in managing user-generated suicide and self-harm content, in conjunction with academics, experts and individuals with lived experience.²⁹ This includes processes for removing detailed information on suicide/self-harm methods, turning off algorithms that push harmful content related to suicide/self-harm, using age and sensitivity warnings, prioritising and promoting positive and helpful content, and effective moderation processes.

Samaritans Ireland believes that the prevalence and placement of harmful online content should be explicitly identified as a key risk of harm that registered service providers should be aware of and measures should be included in the codes both to identify instances of inappropriate display or inappropriate prevalence of content with a risk of harm.

Algorithms which select content for display to a specific user must be developed with an ethical attitude to user behaviour which seeks to minimise compulsive or prolific consumption of difficult content. Research conducted between Samaritans Ireland and Ulster University has shown small changes in service operation, or brief interruptions, can break the cyclical tendencies, and overall, positively impact the behaviours of service users who may have otherwise continuously displayed concerning relationships with the service.

We are aware of algorithms to deliver content on the basis of use / clicks / reach which can result in inappropriate dissemination of content and would recommend the adoption of policies allowing for the moderation or revision of these algorithms to reduce 'doom scrolling' and encourage help seeking without inhibiting individuals' rights to view public content.

²⁹ Samaritans industry guidelines for managing self-harm and suicide content <https://www.samaritans.org/about-samaritans/research-policy/internet-suicide/guidelines-tech-industry/>

Question 19: How do you think that cooperation with other regulators and bodies can help us to implement the Code for VSPS

In 2019, following the death of Molly Russell³⁰, our Samaritans Central Charity (SCC) colleagues in Great Britain, in collaboration with the UK government and some of the largest tech platforms, established an Online Excellence Programme with the aim of promoting good practice around self-harm and suicide content online.

This includes an advisory service for professionals and platforms dealing with self-harm and suicide content online, a published best-practice guidance document³¹ for platforms hosting user-generated self-harm and suicide content, a programme of research to better understand the risks and benefits for users accessing this material, and online user resources to support individuals to talk about suicide and self-harm safely online.

A similar joint process could be followed in Ireland – using a co-design/co-developing approach to design ensures Codes can be understood and enacted effectively for all involved.

Question 20: What approach do you think we should take in the Code to address feeds which cause harm because of the aggregate impact of the content they provide access to? Are there current practices which you consider to be best practice in this regard?

Samaritans Ireland has gained valuable insights from our Online Excellence Programme in the UK as to how platform design, systems and processes can be shaped to enhance the safety of their users, and have developed guidelines for the tech industry in managing user-generated suicide and self-harm content, in conjunction with academics, experts and individuals with lived experience.³² This includes processes for removing detailed information on suicide/self-harm methods, turning off algorithms that push harmful content related to suicide/self-harm, using age and sensitivity warnings, prioritising and promoting positive and helpful content, and effective moderation processes.

However, some of the most harmful suicide and self-harm content exists on smaller platforms. A recent systematic review looking at the impact of suicide and self-harm-related videos and photographs found that potentially harmful content massed on sites with poor moderation and anonymity.³³

³⁰ <https://www.bbc.com/news/av/uk-50186418>

³¹ Samaritans (2020) Samaritans' industry guidelines: Guidelines for sites and platforms hosting user-generated content

³² Samaritans industry guidelines for managing self-harm and suicide content

³³ Marchant, Amanda, Keith Hawton, Lauren Burns, Anne Stewart, and Ann John. Impact of Web-Based Sharing and Viewing of Self-Harm–Related Videos and Photographs on Young People: Systematic Review. *Journal of medical internet research* 23, no. 3 (2021)

Through Samaritans' Online Excellence Programme and Advisory Service, our SCC colleagues regularly receive emails from members of the public concerned about smaller platforms, including bereaved parents whose children have accessed these platforms prior to dying by suicide and practitioners concerned that children as young as 12 are accessing these spaces.

There is a need to protect users from online harms irrespective of the origin hosting platform. Safety codes must ensure all platforms have a duty to protect their users beyond requiring them to confirm they are over 18, and to adequately moderate the content on their platforms, to help ensure they also adapt their platform design, systems, and processes so that risk of harm is minimised.

Question 22: What compliance monitoring and reports arrangements should we include in the Code?

In addition to monitoring and reporting the speed, accuracy, and human level of involvement in removing and reducing self-harm and suicide content online, we also think it is important that services are held accountable for the mental wellbeing of their staff – especially those continuously exposed to distressing content. To ensure compliance with other areas of the online safety code, it is critical that moderators are able to operate at full capacity and effectively remove/reduce harmful or potentially harmful online content.

VSPS providers should be requested to appear before the Online Safety Commissioner and/or relevant committee to report on their compliance on an annual basis. This will help ensure their moderation standards are fit for purpose and that the providers are appropriately managing the balance between human-to-AI moderation ratio, while also ensuring their human moderators receive high quality training and support.

Samaritans Ireland believes the risks to moderators wellbeing is directly related to the reduction in quality of moderation and should be explicitly addressed within the codes. Outlining compliance monitoring and reporting of this nature in the Online Safety Codes is a key way to monitor internet safety. The monitoring/report should include specific measures for platforms to ensure the good mental health and wellbeing of people who review/moderate potentially harmful content including things like mandatory reporting on support measures in place for any persons who review, categorise, edit and/or remove harmful or potentially harmful content including things like formal/informal debrief, job rotation, breaks, training, and professionals supports as needed.

From our research and experiences with our own volunteers, we know that exposure to self-harm and suicide content, particularly over an extended period, can negatively affect mental wellbeing. We have developed robust internal support mechanisms for our volunteers to limit harm and enable them to operate at their highest, healthiest capacity thereby also better serving the needs of vulnerable people.

Samaritans Ireland knows it is of the upmost importance that everyone be given the opportunity to ensure their mental health and wellbeing is looked after – this allows individuals to be happier, healthier, and therefore more equipped to successfully do their job. Any programme to manage online harms must take account of the health and wellbeing of content moderators both to protect and support a specific vulnerable or ‘at risk’ group but also to improve the standard of the moderation itself, avoiding relevant personnel being desensitised or burnt out and thereby less able to appropriately moderate making the internet less safety.

Prepared by Samaritans Ireland

4 September 2023

Contact: Sarah O’Toole, Executive Director for Samaritans Ireland.
Email: [REDACTED]
Tel: [REDACTED]

SUBMISSION - Call For Inputs: Developing Ireland's First Binding Online Safety Code for Video-Sharing Platform Services

Link to call: https://www.cnam.ie/wp-content/uploads/2023/07/CallForInputs_vFinal.pdf

Organisation: Eurochild

Description: Eurochild is a network organisation with almost 200 members in 41 countries. We are striving for a society where all children and young people grow up happy, healthy, confident and respected as individuals in their own right. We aim to bring about positive changes in the lives of children, in particular those affected by poverty and disadvantage. We have been working closely with the European Union and at national level with our members to foster regulatory and policy opportunities to enhance children's safety and the fulfilment of their rights online.

Contact(s):

Mieke Schuurman, Director of Child Rights and Capacity Building
[REDACTED]

Fabiola Bas Palomares, Policy & Advocacy Officer on online safety
[REDACTED]

Part I: Online harms

Question 1: What do you think our main priorities and objectives should be in the first binding Online Safety Code for VSPS? What are the main online harms you would like to see it address and why?

The main objective of the code should be to enable video-sharing platforms to take responsibility in preventing harm to children using their services. For this, the code should cover a broad spectrum of online harms and go beyond harmful content. The code should also cover harmful contact and conduct and the contract risks that arise from children using video-sharing platforms. This means the scope of the code should include:

- Harmful content: including sexual & violent videos; videos that promote hate speech; content fostering self-harm or suicide, etc.
- Harmful contact: chat/comment/post functions being used by adults or other children to lure a child user into harmful activities (i.e., grooming)
- Harmful conduct: children writing or creating hateful materials about other children, inciting racism or hate, posting or distributing sexual images, etc.
- Contract risks: ensure the terms & conditions of video-sharing services do not bind the child user in ways that may be unfair or exploitative, or which pose security or safety or privacy risks he/she might not be aware of.

Question 2: What types of online harms do you think should attract the most stringent risk mitigation measures by VSPS? How could we evaluate the impact of different types of harms e.g. severity, speed at which harm may be caused? Is there a way of classifying harmful content that you consider it would be useful for us to use?

The code should follow the classification of harm used in the response to Question 1, because of two reasons. Firstly, it is based on risk, which will facilitate the risk assessment exercise. Second, it allows for much more multidisciplinary. Instead of using narrow categories such as “self-harm” or “sexual harm”, the classification above allows you to discover the different manifestation of the harm thus facilitating the identification of the root of the harm. This is key for effective risk mitigation measures. For example, the broader category of ‘sexual harm’ can unfold into ‘sexual coercion’ when it is due to harmful contact or into ‘child pornography’ when it is to harmful content. These two require very different responses. Therefore, this classification can enrich the risk assessment and mitigation exercise considerably.

Following this classification – Because of the nature of the service provided by video-sharing platforms, the focus should inevitably be put in harmful content. However, there is some history and much good practice of such platforms filtering, detecting or removing harmful content. Therefore, the code should pay special attention to the harms from contact (i.e., chat moderation tools, detecting hate speech or violence against children in public chat rooms and comments, etc.).

Question 3: Do you have reports, academic studies or other relevant independent research that would support your views? If you do, please share them with us with links to relevant reports, studies or research.

- Livingstone, S., & Stoilova, M. (2021), ‘[The 4Cs: Classifying Online Risk to Children](#)’
- Stoilova, M., Rahali, M. & Livingstone, S (2023), ‘[Classifying and responding to online risk to children: Good practice guide.](#)’

Section II: Overall approach to the code

Question 4: What approach do you think we should take to the level of detail in the Code? What role could non-binding guidance play in supplementing the Code?

Having regarded the three options foreseen in the call, and taking into account this would become a binding code but only upon subscription by the providers, the code should pursue a mixed approach (option 3). The Code could consist of over-arching principles, specific commitments and KPIs for monitoring compliance. Setting the high-level principles will allow for innovation in terms of good practices and will ensure the code remains tech neutral and future proof; while the specific commitments will ensure the code is actionable and measurable.

Question 5: What do you think would be the most effective structure for the Code? What are the most important factors we should consider when we decide how to structure the Code?

As a binding tool and to maximize compliance, the code should be structured in the terms and items of the AVMSD, for example the 10 measures provided by Article 28b.3, and from the DSA where possible. These obligations could be grouped into high-level principles and further concreted into actionable commitments. However, it is very important to remain ambitious and aim to include other commitments that advance online child protection, for example by harnessing good practice that falls outside the scope of the DSA and the AVMSD, in particular innovative solutions fostering safety-by-design.

Question 6: How should we design the Code to minimise the potential for conflict and maximise the potential for synergies in how platforms comply with it and the DSA?

The code is a good opportunity to operationalise the DSA provisions for video-sharing providers, especially in relation to Article 28 but also 14.3 (Terms and Conditions). For this, the code could delve into the use & design of recommender systems, predominant in video-sharing services, and [which can have devastating effect on children](#). For instance, the code could include provisions to carry out child rights impact assessments when developing recommender systems or algorithms. However, it is very important that the code remains ambitious and effectively enforces and extends some provisions from the DSA, where possible. It should not in any case fail to comply with the current regulation at EU level, most notably the AVMD and the DSA. For example by extending measures under the DSA for VLOPs to all video-sharing service providers subject to the code [Articles 34.1 (risk assessment) & 35.1 (risk mitigation)]. The ‘trusted flaggers’ included in the DSA could be extended under this code to all video-sharing providers through the code, as well as some responsibilities from the DSA Coordinator. In parallel to the Terrorist Content Online Regulation (TCOR), The Code of practice on disinformation is a useful example.

Section III: Content connected to video content

Question 7: To what extent, if at all, should the Code require VSPS providers to take measures to address content connected to video content?

Question addressed already in Q2.

Section IV: Measures to be taken by Video-Sharing Platforms

Question 8: How should we ask VSPS providers to introduce a feature that allows users to declare when videos contain advertising or other type of commercial communications? Should the Code include specific requirements about the form in which the declaration should take? What current examples are there that you regard as best practice?

The code must remain ambitious and take the opportunity to set good practice and include specific requirements when possible. In this case, the code should require such a feature to be age-appropriate, child-friendly and in a child-friendly language. The language used to and the way the tool to declare commercial videos should be adaptable to the age of children, to ensure the tool is effective in preventing unwanted effects on children. This means it should make sure the child, regardless their age, understand what “contain commercial communications” means and is able to answer meaningfully (i.e., the options provided by the feature are understandable to children). Ideally, such features should be developed with (or at least tested with) children.

Question 9: How should we ask VSPS providers to introduce and design a flagging mechanism in the Code? How can we ensure that VSPS providers introduce the mechanism in a user-friendly and transparent way? How should we ask VSP Providers to report the decisions they’ve made on content after it has been flagged? To what extent should we align the Code with similar provisions on flagging in the DSA?

Although it is true in the case of reporting it is difficult to set common requirements, the code should set some common minimum standards. From what we have been hearing from children, they often tend to block content instead of reporting when they feel at risk. This is due to several causes, but two of the most raised are the complexity of the process and the lack of follow-up from the service provider. The code should incentivise the simplification of reporting features (i.e.,

reduce the options to categorize the reported content, shorten the process by asking less details) and the use of child-friendly language. To ensure these systems meet children's needs, the code should encourage platforms to involve children themselves in the design of such reporting mechanisms.

Moreover, it should incentivise good practice where service providers ensure follow-up to children who report with information on what happened to the reported content/user and what to do if the child encounters a similar situation (i.e. the user created another account or the content appears again). In cases of reports of cases of extreme violence (i.e., child sexual abuse, bullying) the providers should be encouraged to provide the child with information on where to seek help.

Question 10: What requirements should the Code include about age verification and age assurance? What sort of content should be shown by default to users who are logged out or in private browsing mode and whose age cannot be verified or assured? What evidence is there about the effectiveness of age estimation techniques? What current practices do you regard as best practice? Where accounts are not age verified should default privacy settings be used, should content default to universal content and should contact by others be more limited?

The objective of the code in this regard should be to decrease access (intentional or by accident) of children to adult services and content. The European Digital Identity Wallets does not seem to work for children. While Member States have the option to issue existing eIDAS credentials to children, very few already do so, and then usually only to older teenagers. In addition, there will be usability issues for younger kids which may make providing only age-appropriate content harder to do. Moreover, restricting access to age-restricted content to eIDAS holders only would immediately exclude or delay a large number of children, especially non-EU, migrant and undocumented children.

Eurochild is part of the Management Board of euConsent, which aims to deliver a pan-European, open-system, secure and certified interoperable age verification and parental consent to access Information Society Services. When working with children, their level of trust towards ID-based age verification systems is alarmingly low. From this, we believe that age verification requirements have to be homogenous across the sector and lead to a one-stop verification system. This way, the system becomes more accessible and trustworthy to children.

Regardless of the specificities of the system chosen (ID, fingerprint, a password, facial recognition) there are two principles that must be encouraged by this code. First, it has to be child-friendly and accessible to all ages that the service targets. Second, it must be privacy-preserving (minimizing the personal data collected and stored by platforms when verifying age, i.e. name, address, or date of birth).

The code needs to recall, however, that age verification measures must be complemented with age assurance measures. The best interests of the child plays a central role here. However, since current age verification methods do not work for all children (i.e., migrant or undocumented children, younger children, children with disabilities, etc.), robust reporting mechanisms must be in place for children. Children shall be protected from harm but also empowered as digital citizens. Age assurance will promote the right to participation while ensuring children can foster all the opportunities the digital environment offers, instead of keeping children from using a particular service.

[Sidenote: from talking to the children, we have learned that in services with self-declaration age verification, children not only declare being older to be able to access, but also some times to protect themselves so that other users do not target them for being children]

Question 11: What requirements should the Code have in relation to content rating? What do you consider to be current best practice? What experiences have you had using content rating systems on platforms and do you think they have been effective? What steps could we ask VSPS to take to ensure content is rated accurately by users?

Question 12: What requirements should the Code have in relation to parental control features? How can we ensure that VSPS providers introduce the mechanism in a user-friendly and transparent way? Can you point to any existing example of best practices in this area? Should parental controls be ‘turned-on’ by default for accounts of minors or where age is not verified?

Parental controls have proven to be insufficient to ensure child safety online. However, the code could be an opportunity to improve some aspects; for example, the interoperability of parental controls among video-sharing platforms. However, the code should focus on expanding ‘safety-by-design’ features that prevent harm as opposed to blocking access, including turning on high privacy settings and parental controls by default when the user registers or is identified as a child. Both the child and the parent should be informed about this by the service to ensure transparency and empower children’s agency online.

Question 13: What requirements should the Code contain to ensure that VSPS provide for effective media literacy measures and tools?

Question 14: How should we ask VSPS providers to address online harms in their terms and conditions in the Code, including the harms addressed under Article 28b? How should key aspects of terms and conditions be brought to users’ attention? What examples are there of best practice in relation to terms and conditions including content moderation policies and guidelines?

The code should pay attention to terms & conditions (T&C) when addressing contract risks and, consequently, other types of harm online. It should set minimum standards for content moderation content policies and sufficient consequences for those who break them. Illegal content should not be tolerated and the mechanisms for effective detection and removal of such content should be tightened. It should include prohibition on any content that exploits the vulnerabilities of children. It is key that the list of requirements address all the harms mentioned above: content, contact, conduct and contract (including violence, discrimination, disinformation, sexual harm, commercial harm, manipulation, etc.). Above those minimums, it could encourage the providers to enable children to adjust the content they want to have access (for example, by blocking certain key words or tags).

Every platform should be required to have a child-friendly version of their T&C to ensure children can give their informed consent. Some [good practice on age-appropriate presentation of T&C](#) was gathered by 5Rights Foundation.

Question 15: How should we ask VSPS providers to address content moderation in the Code? Are there any current practices which you consider to be best practice? How should we address automated content detection and moderation in the Code?

Regarding the accuracy of content moderation, the code should establish reporting requirements for video-sharing platforms. This would incentivize the companies to improve their content moderation practices iteratively and facilitate innovation.

However, we know content moderation and self-reporting, especially among children, are not enough and lead to low levels of removal of harmful content. Therefore, we encourage the code to address automated content detection obligations for illegal content (i.e., child sexual abuse). The ‘trusted flaggers’ included in the DSA could be extended under this code to all video-sharing providers through the code, as well as some responsibilities from the DSA Coordinator.

Question 16: What requirements should the Code include about procedures for complaint-handling and resolution, including out-of-court redress or alternative-dispute resolution processes? To what extent should these requirements align with similar requirements in the DSA? What current practices could be regarded as best practice? How frequently should VSPS providers be obliged to report to the Commission on their complaint handling systems and what should those reports contain? Should there be a maximum time-period for VSPS providers to handle user complaints and if so, what should that period be?

Section V: Additional measures

Question 17: What approach do you think the Code should take to ensuring that the safety measures we ask VSPS providers to take are accessible to people with disabilities?

Question 18: What approach do you think the Code should take to risk assessments and safety by design? Are there any examples you can point us towards which you consider to be best practice?

As highlighted before, the code should complement and expand the DSA but avoid overlaps as much as possible. A possibility for expanding such obligations would be to require VSPS providers designated as VLOPs to carry out a similar risk assessment than that of the DSA but applied to the content in scope of this code, integrated as much as possible with the requirements of the DSA the provider might have to comply with. For VSPS who are not VLOPs, a reduced version of such assessment could be required. These risk assessments should include or be complemented with a children’s rights impact assessment.

Regarding safety-by-design, it is strongly recommended that the code to focuses on some minimum requirements on safety and privacy by design (some of which have already been mentioned in other questions). For these measures to be truly effective, children must be involved and consulted during the design phase of new features. The code could contain some requirement for video-sharing platforms used by or targeted at children to consult children about their safety features (i.e., reporting mechanisms, privacy by default settings, parental controls, content moderation, etc.). This could be done as part of the child-rights impact assessment.

Question 19: How do you think that cooperation with other regulators and bodies can help us to implement the Code for VSPS?

Question 20: What approach do you think we should take in the Code to address feeds which cause harm because of the aggregate impact of the content they provide access to? Are there current practices which you consider to be best practice in this regard?

Companies must design digital services that cater for vulnerabilities, needs, and rights of children and young people by default. Specific rules for algorithms and recommender systems that exploit the vulnerabilities of children should be included in the code, and must apply to all video-sharing platforms that children are likely to access in reality, not just services specifically targeted at them. For example, platforms should make it easier for children to be aware of the time spent interacting within a service, and should incentivise them to take a break from time to time. The

platforms and the algorithms they use should be designed in a way that protect children's identity and automatically block unwanted contact or content.

Question 22: What compliance monitoring and reporting arrangements should we include in the Code?

Question 21: Do you have any views on how requirements for commercial content arranged by a VSPS provider itself should be reflected in the Code?

Question 23: Should the Code have a transition period or transition periods for specific issues? Which areas touched on in this Call for Inputs may VSPS providers require time to transition the most? What time frame would be reasonable for a transition period?

Call for inputs: Online Safety Code for Video-Sharing Platform Services

Australia's eSafety Commissioner (eSafety) welcomes Coimisiún na Meán's proposal to establish a binding online safety code for Video-Sharing Platform Services (VSPS), and the opportunity to contribute to this consultation process.

eSafety is pleased to see plans to introduce stronger regulation of VSPS to better protect users from the broad range of significant online harms that can occur on these services. We hope that the development of your code will provide better protection to users, address harmful content, promote shared responsibility of user safety, promote human rights, and encourage providers to adopt a Safety by Design approach to product development and deployment.

Recognising the close connections between eSafety and Coimisiún na Meán, bilaterally and through a Global Online Safety Regulators Network, we wish to take this opportunity to provide information about Australia's online safety regulatory framework as set out in the *Online Safety Act 2021*, including the Basic Online Safety Expectations, and industry codes and standards. We have also provided links to range of resources that could be considered when drafting your code.

Overview of the eSafety Commissioner model

The eSafety Commissioner (eSafety) is Australia's independent regulator for online safety. eSafety promotes online safety for all Australians, leads online safety efforts across Australian Government departments and agencies, and works with online safety stakeholders around the world to extend our impact across borders. Established in 2015, our mission is to help safeguard Australians at risk from online harms and to promote safer, more positive online experiences.

eSafety has a broad remit and powerful combination of functions, which enables us to address online safety in a multifaceted and holistic way. We do this through:

- **Prevention** – by providing evidence-based resources and programs;
- **Protection** – by operating regulatory schemes and investigating abuse;
- **Proactive and systemic change** – by identifying emerging risks, ensuring industry minimises harm, and collaborating to lift overall standards; and,
- **Partnerships** – by working collaboratively with domestic and international partners across government, industry, and civil society to amplify our reach and impact.

Australia's regulatory framework

The Australian Government passed the Online Safety Act 2021 (the Act) to better equip eSafety to prevent and address online harms. The Act enhances eSafety's regulatory schemes for dealing with the cyberbullying of children, adult cyber abuse, image-based abuse, and illegal or restricted online content. Our [Regulatory guidance](#) documents explain how eSafety implements each of the regulatory schemes included in the Act.

Illegal and restricted content ranges from material showing the sexual abuse of children or acts of terrorism (defined as class 1 material), through to material that should not be accessed by children, such as simulated sexual activity, detailed nudity or high impact violence (defined as class 2 material). Class 1 and class 2 material are defined by reference to the [National Classification Scheme](#).

eSafety also has [Abhorrent Violent Conduct Powers](#) to prevent Australian internet users from accessing material that promotes, incites, instructs in or depicts abhorrent violent conduct, which is defined as material that, promotes, incites, instructs, or depicts abhorrent violent conduct, such as terrorism, murder or attempted murder, torture, kidnapping, and rape.

The Act places more stringent requirements in relation to ‘class 1 material’, recognising the significant harm that is caused through its production, distribution and consumption. The Act also regulates online services’ systems and processes through two regulatory schemes: The Basic Online Safety Expectations and mandatory industry codes and standards, which focus on illegal and restricted content.

The Basic Online Safety Expectations

The Act allows for the creation of a set of ‘Basic Online Safety Expectations’ (the Expectations) for online service providers. The Online Safety (Basic Online Safety Expectations) Determination 2022 was made by the former Minister for Communications, Urban Infrastructure, Cities and the Arts on 20 January 2022, setting out a series of expectations that services will take reasonable steps to:

- make sure end-users can use the service in a safe manner,
- have, and enforce, terms of use, policies and procedures in relation to the safety of end-users, and
- provide clear and readily identifiable ways for end-users to report specific forms of harmful content or behaviour to the service.

Under the Act, eSafety can require online services to report on the extent to which they are complying with the Expectations. This provides greater transparency around their safety features, policies and practices, noting that services’ transparency to date has been selective and uneven.

In deciding whether to issue a notice that requires an online service to submit a report, eSafety is required by the Act to consider several factors. These include the number of complaints received about material on the service, any safety deficiencies, and other factors that eSafety has said it will have regard to such as the existing transparency of a service. The first non-periodic reporting notices were given on 29 August 2022 to Apple, Meta, WhatsApp, Microsoft, Skype, Omegle and Snap on how these providers are addressing child sexual exploitation and abuse (CSEA). eSafety published a [report](#) in December 2022 summarising the findings.

In February 2023, a further five notices were given to Google, Twitter/X, Twitch, TikTok and Discord, also focussing on CSEA as well as sexual extortion and the safety of recommender systems. A notice was also given to Twitter/X in June 2023 on how Twitter is addressing and enforcing its policies around online hate. Safety intends to publish reports of the information obtained through these notices. Reports from this round of notices will be published. Further information about the notices can be found [here](#).

This year, we will continue to utilise these reporting powers to require information from different services and on other harms. eSafety also has powers to require periodic reporting to track key issues and metrics over time.

Industry Codes and Standards

The Act provides for industry bodies or associations to develop a new code, or set of codes, to regulate certain types of online material, and for eSafety to register the codes if they meet the statutory requirements. If a code does not meet the requirements, then eSafety can develop an industry standard for that section of the online industry instead.

Industry Codes and standards apply to the participants of eight key sections of the online industry, including providers of social media, messaging, search engine and app distribution services, as well as websites and online file storage, internet and hosting service providers, manufacturers and suppliers of equipment used to access online services, and those that install and maintain the equipment.

To assist the online industry to develop codes, eSafety issued a [position paper](#) in September 2021. The paper set out 11 policy positions regarding the substance, design, development and administration of industry codes, as well as eSafety's preferred outcomes-based model for the codes. It paper outlined a set of drafting principles that should underpin code development. These guiding principles were that codes should be:

- consistent (where there are multiple codes)
- clear
- meaningful
- implementable
- measurable
- proportionate
- respectful of rights, and
- striking a balance between safety, privacy and security.

An outcomes-based approach can provide services with a common set of objectives and outcomes, while granting the flexibility to implement measures to meet those objectives and outcomes that are most suited to their business models and technologies. In addition, and to ensure high-risk industry participants can be held to account, eSafety considered that an outcomes-based model should be supported by clearly defined compliance measures for each outcome, applying to industry participants whose services and devices present the greatest risk of online harm.

Industry associations adopted a two-phase approach to code development. The first phase focused on class 1 material, including child sexual exploitation material and pro-terror material.

In June 2023, after nearly two years of negotiations, the Commissioner registered five codes for class 1 material. The codes apply to:

- Social Media Services
- App Distribution Services
- Hosting Services
- Internet Carriage Services, and
- Equipment providers.

Coimisiún na Meán may find the registered industry codes to be a useful reference point. Copies of the industry codes are available on the [Register of industry codes and industry standards for online safety](#).

These codes will come into effect on 16 December 2023 and will require the online industry to take proactive steps to protect Australians from being exposed to this content. This includes a

requirement for major Social Media Services to proactively detect and remove known (i.e. previously identified and verified) child sexual abuse material and known pro-terror material.

The Commissioner decided not to register two industry codes because they did not provide appropriate community safeguards to users in Australia. These draft codes covered Relevant Electronic Services (including peer-to-peer communication such as instant messaging, dating sites and online games) and Designated Internet Services (including apps, websites and online file and photo storage services). eSafety is now developing industry standards for these industry sections.

The Commissioner reserved the decision on the Search Engine Services code because the code did not adequately capture new risks emerging from the integration of generative artificial intelligence features. The Commissioner is currently considering a revised version of the code submitted by industry associations.

eSafety is currently mapping out the most appropriate compliance and enforcement model for the first phase of the codes and will provide guidance in due course. We would be happy to discuss approach to compliance monitoring and reporting with Coimisiún na Meán as it develops its code.

The second phase of codes development will focus on class 2 material such as online pornography, and will commence after Phase 1 has been completed.

Age verification

In June 2021, the Australian Government requested that eSafety develop a roadmap for a mandatory age verification regime relating to online pornography, which aims to mitigate harms associated with children and young people's access to online pornography. In developing the Roadmap, eSafety examined a range of age assurance measures, which was presented to government in March 2023. You can read the Roadmap and background report [here](#).

After almost two years of careful consultation and analysis, eSafety reached the position that there is no regulatory or technical silver bullet when it comes to protecting children from harmful, age-inappropriate content. Instead, a holistic approach is needed that respects and affirms a range of human rights and promotes the best interests of the child through technological and educational measures.

As outlined in the Roadmap, factors such as privacy, security, data minimisation, equity and inclusion, choice, trust, accuracy and the impact on competition will be key in assessing the efficacy, proportionality and feasibility of any age assurance regulation. On this basis, we recommended testing age-assurance technologies in an Australian context through a pilot project, while continuing to observe international technical and regulatory developments, before moving to mandating technical measures.

In its response, the Australian Government acknowledges concerns about the effectiveness, privacy and security of some age-assurance technologies. The response also recognised the nascent state of the age-assurance industry, and the comprehensive work already being done across government to address intersectional issues identified in the roadmap.

Recognising that the Codes require industry to consider how to limit children's access to 'Class 2 material', which includes online pornography, the government deferred the decision to pilot age verification technologies following the conclusion of the industry codes/standards process for 'Class 2 material'.

Under the Act, the Expectations also encourage industry to take a range of steps to ensure that children's experiences on online services and platforms are not harmful. This includes expectations regarding children's access to certain material and ensuring safe use of a service through mechanisms such as high privacy and safety default settings. eSafety is able to compel reports from providers of what they are doing to meet Expectations under the Act to improve transparency and accountability, including on the steps to prevent children's access to pornography.

Safety by Design and other initiatives

eSafety's Safety by Design initiative is a voluntary, non-regulatory measure. It provides a framework for industry action in terms of the design, development and deployment of online services. It is an initiative that puts user safety and rights at the centre of design and development of online products and services. eSafety has developed a range of initiatives and tools to support industry embed Safety by Design within product development and deployment. These include:

- a set of principles that position user safety as a fundamental design consideration
- interactive assessment tools for enterprise and start up technology companies
- resources for investors and financial entities
- engagement with the tertiary education sector to embed Safety by Design into curricula around the world.

The tools provide guidance to industry on a range of online safety issues, including measures that online services can implement to ensure users can understand and access reporting and complaints mechanisms, improved content moderation, and responding to terms of service breaches.

We welcome the international uptake of the principles and use of the interactive assessment tools to assist industry to prioritise user safety.

In addition to initiatives like Safety by Design, eSafety is focused on preventing a broad range of online harms, through education and awareness raising. Recently, eSafety supported the World Economic Forum's Global Coalition for Digital Safety in their development of the Typology of Online Harms (the Typology), which seeks to provide a foundational common language for key online harms terminology.

The Typology forms part of the Toolkit for Digital Safety Design Interventions and will promote shared understandings, support cross-jurisdictional dialogue and multi-stakeholder discussions, and ultimately advance online safety through cohesion and harmonisation. The Typology also seeks to provide a comprehensive framework that supports organisations in their understanding of the interplay between harms and the categorisation of risk. The Typology could be a useful tool for regulators in the development of codes and standards as it provides a foundation of understanding that considers the interplay between online harms and new technologies.

Collaboration

Recognising the importance of collaboration, in March 2022, the Australian Communications and Media Authority, the Australian Competition and Consumer Commission, the Office of the Australian Information Commissioner and eSafety formalised existing arrangements to establish the Digital Platforms Regulation Forum (DP-REG).

Through DP-REG, members share information about, and collaborate on, cross-cutting issues and activities on the regulation of digital platforms. This includes consideration of how competition,

consumer protection, privacy, online safety, information integrity and data issues intersect - which facilitates greater understanding of our respective regulatory agendas - and the ability to identify and anticipate opportunities for coordination and tensions that may arise.

We also recognise that we are part of an international regulatory ecosystem, and we welcome new regulators and novel approaches being tested to improve online safety globally. As you are aware, eSafety is working with international partners to share what we learn - and to learn from their planned approaches, through the Global Online Safety Regulators Network and bilaterally.

We are delighted to be collaborating with Coimisiún na Meán, because we know that leveraging the collective insights and expertise of other regulators, both domestically and internationally, will make sure that best practice continues to evolve, and we are able to tackle emergent threats.

As always, we would welcome future opportunities to meet with you to discuss the contents of this submission in more detail. We look forward to future opportunities for engagement and collaboration.

TikTok Technology Limited

Online Safety – Developing Ireland’s First Binding Online Safety Code for Video-Sharing Platform Services

TikTok’s commitment to online safety

TikTok welcomes the opportunity to make submissions in response to Coimisiún na Meán’s (the **Commission**) call for inputs on the development of Ireland’s first Online Safety Code (the **Code**) (the **Call for Inputs**).

At TikTok, we are focused on providing our community with a platform that offers a joyful, creative, and above all, safe experience. TikTok is a diverse, global community fueled by creative expression. We work to maintain an environment where everyone feels safe and welcome to create videos, find community, and be entertained. We believe that feeling safe is essential to feeling comfortable and expressing yourself authentically. To this end, TikTok has taken a safety by design approach to protect our users’ safety and well-being.

We do this by countering harmful misinformation, tackling deceptive behaviour and otherwise taking measures to prevent harm to our community or society at large. Our [Community Guidelines](#) and [Terms of Service](#) define a set of norms and common code of conduct for TikTok; they provide guidance on what is and is not allowed to help maintain a welcoming space.

We are committed to being transparent about how our policies are enforced, because it helps build trust with our community and holds us accountable. We publish Transparency Reports to provide visibility into the volume and nature of content removed for violating our Community Guidelines or Terms of Service.

There is no finish line when it comes to protecting the TikTok community. We work each day to learn, adapt, and strengthen our policies and practices to keep our community safe. We take our responsibility to protect our community incredibly seriously. It is the most important work we do, and we will continue to innovate to keep our community safe.

TikTok’s commitment to engaging in the consultation for the Online Safety Code

We understand that the Code will focus on video-sharing platform services (**VSPSs**) and note that the Commission designated VSPSs as a category of services under the Online Safety and Media Regulation Act 2022 (**OSMR**) on 14 August 2023.

TikTok aims to support the Commission in the development of the Code by providing its views on aspects of how the proposed Code should regulate VSPSs in order to ensure the Code is effective, practical, proportionate, and legally robust in line with the objectives to be achieved. To this end, we have considered the

questions raised in the Call for Inputs and have set out our responses to those questions where we believe we can assist the Commission in its task.

We have also outlined below four key overarching themes arising from TikTok's responses to the questions raised in the Call for Inputs which we would strongly encourage the Commission to take into account in the framing and drafting of the Code:

1. The need for a principles-based (rather than prescriptive) approach to the Code

We note that the Commission has indicated its current intention to develop one Code for VSPSs. As rightly noted in the Call for Inputs, VSPSs vary in terms of their users, size and the kinds of content they make available. In recognition of this and the fact that one size will undoubtedly not fit all, the Code should work proportionally and fairly for the full variety of different video sharing services it will regulate. In order to achieve this, we believe that a principled and risk-based approach to regulation via the Code is the most appropriate approach. This would involve the Commission setting out at a high-level the categories of harm that VSPS providers must address and obliging them to take appropriate and proportionate measures to reduce the risk of harm in general terms, supplemented by non-binding Commission guidance. Such guidance could be used to provide the Commission's view on the factors that can be taken into account when assessing compliance with the principles of the Code. This approach would best ensure that the Code is future-proofed, adaptable and aligned with the requirement of the revised AVMS Directive (**AVMSD**). This is also an approach which has been endorsed in other European jurisdictions, including in the Netherlands (which simply requires VSPSs to adopt a code of conduct which provides for the measures laid down in the AVMSD) and the United Kingdom (see further paragraph 4 below). This will also best ensure that the measures VSPSs are required to take are practicable and proportionate, taking into account their size and nature.¹ Such an approach also fits fully within the overall context of Article 28b(3) of the AVMSD which expressly provides that the measures to be taken must be those which are "appropriate" - and it may, e.g., be the case that not all measures are appropriate for all VSPSs. In our view, an overly prescriptive Code risks unduly constraining VSPSs in their approach to online safety and may fail to take account of developing/changing platform technologies, resulting in the Code being less effective in achieving the legislative goals. In addition, there is also a potential risk that imposing prescriptive obligations on VSPSs may run the risk of discouraging them from implementing new protective innovations for online safety.

2. The Code should not conflict with other EU regulatory regimes

As acknowledged in the Call for Inputs, the subject area of the Code is complex and is governed by different legal instruments, many of which seek to achieve the important goal of harmonisation in the online safety space. We would therefore encourage the Commission to have due regard to the existing obligations VSPSs are or may be subject to under other regulatory regimes in the European Union (**EU**), importantly the Digital Services Act (**DSA**), when developing the Code. As the Commission is aware, VSPSs will have to comply with obligations for online platforms under the DSA, with an increased set of DSA obligations for VLOPs.² As

¹ Article 28b(3) of the AVMSD states that "...Member States shall ensure that all video-sharing platform providers under their jurisdiction apply such measures. Those measures shall be practicable and proportionate, taking into account the size of the video-sharing platform service and the nature of the service that is provided."

² As the Commission will be aware, TikTok has been designated by the European Commission as a provider of a very large online platform (**VLOP**) under the DSA.

such, it will be crucial that the Code does not undermine the harmonised approach to online safety mandated by the DSA. In order to avoid duplicative regulatory requirements and to seek to prevent any conflicts with the DSA, we consider that the Code should build on the obligations already applicable to VSPS and should not contain additional national requirements relating to matters falling within the scope of the DSA, as this would be contrary to the harmonised approach mandated by the DSA.³

3. The Code should not seek to go beyond the requirements of the AVMSD

In order to reduce the risk of the Code conflicting with other regulatory regimes, we believe the Code should focus on transposing Article 28b of the AVMSD. To the extent that the Code goes beyond the requirements of the AVMSD, we are conscious that this would risk further delaying the transposition of the AVMSD in Ireland, as it will increase the likelihood of issues being raised when the Code is notified to the European Commission via the Technical Regulation Information System (TRIS) procedure, thus undermining the goal of creating effective, practical, proportionate and legally robust measures in line with the requirements of the AVMSD.

4. The transposition of the AVMSD in other EU Member States suggests this is also appropriate in Ireland

Focusing the Code on the requirements of the AVMSD is consistent with the transposition of that legislation in a significant number of EU Member States, where national legislation predominantly reflects the requirements outlined in Article 28b(3) of the AVMSD and, in many cases, it appears that the requirements have been transposed verbatim. We would note that one of the key findings of the European Audiovisual Observatory's [publication](#) on the mapping of national rules applicable to VSPS⁴ (the **EAO Publication**) was that the transposition of the revised AVMSD has resulted in the adoption of legislation in other EU Member States which "very much corresponds to the provisions of the revised AVMSD itself"⁵ and that there has not been much further elaboration or introduction of stricter obligations for VSPSs. Accordingly, we consider that the Commission should adopt a similar approach under the Irish regime so as to avoid the patchwork implementation of the AVMSD across different EU jurisdictions.

Indeed, as a VSPS that has effectively been complying with the AVMSD requirements for a number of years, as part of the UK's 'Video Sharing Platform' framework (including working with [Ofcom's VSP Guidance](#)), we hope that we will have particularly useful and relevant feedback.

Finally, we hope that these submissions will assist the Commission in developing a Code which is fit for purpose, clear, workable and legally robust and we look forward to engaging further with the Commission on this issue following the Commission's intended publication of a draft Code later this year.

We thank the Commission for the opportunity to respond to its Call for Inputs.

³ Note that Recital 9 states that "Member States should not adopt or maintain additional national requirements relating to the matters falling within the scope of the DSA, as this would affect the direct and uniform application of the fully harmonised rules under that regime."

⁴ Mapping of national rules applicable to video-sharing platforms: Illegal and harmful content online – 2022 update

⁵ Section 1.2.1 of the EAO Publication

Section 3: Online Harms: What online harms should the Code address?

Question 1:

What do you think our main priorities and objectives should be in the first binding Online Safety Code for VSPS? What are the main online harms you would like to see it address and why?

What do you think our main priorities and objectives should be in the first binding Online Safety Code for VSPS?

The main priority and the key focus of the first Code for VSPSs should be the effective transposition of the AVMSD, in particular of Article 28b. As highlighted in the Call for Inputs, the Commission is responsible for making sure that VSPS providers under the jurisdiction of the Irish State take these measures through online safety codes and by this means intend to complete the transposition of Article 28b in Ireland.

We are of the opinion that the Code should be limited to the transposition of the AVMSD only. To the extent that the Commission wishes to deviate from the AVMSD provisions (including prioritising categories of online harms), this would require very careful consideration. Any clear deviation from the AVMSD through additional national requirements also increases the risk of the Code regulating matters that fall within the scope of other laws, importantly the DSA. Any conflicts with the DSA would delay the implementation of the AVMSD (the transposition of which is currently an urgent priority for the Commission and Ireland) and may also separately require an additional TRIS notification. That delay would also lead to legal uncertainty for VSPSs under the jurisdiction of Ireland. This would undermine the important goals to be achieved by the AVMSD /OSMR.

In the first instance, we believe the Commission should focus on AVMSD transposition at this stage and, separately, monitor and assess how the additional harm areas covered by the OSMR are dealt with by other regulatory regimes for a period before determining what, if any, additional guidance or codes are required in respect of those additional harm categories.

What are the main online harms you would like to see it address and why?

Article 28b of the AVMSD is clear on the categories of harms in respect of which VSPSs must take appropriate measures to ensure the protection of minors and the general public against those categories of harms. Those categories are broadly drafted, with no specific category prioritised over another - i.e. there is no “hierarchy” of harm in the AVMSD. This reflects the fact that what constitutes a “harm” is a subjective assessment that involves evaluation of the context and circumstances in which the actual/potential harm arose. What may constitute “harmful” content also varies from individual to individual, with one particular or category of harm having a more profound effect on one person than on another. In addition, one particular harm may fall within a number of broad categories.

	<p>Due to the subjectivity associated with determining what may or may not amount to harmful online content, and the case-by-case assessment that will need to be made by VSPSs to combat such online harms in the context of their own specific platforms (based on an assessments of risks specific to that platform), it is both correct and appropriate for potential harms to be defined by way of broad categories. We would therefore encourage the Commission to incorporate in the Code the categories of harms as outlined in Article 28b(1) AVMSD, without specifying which category or harm is deemed to be of higher priority or particular focus. “Picking and choosing” which harms should or should not be prioritised and attempting to reframe or redefine these harms in the binding Code may have the unintended consequence of limiting or distorting the application of Article 28b AVMSD.</p> <p>Alternatively (and without prejudice to our position that the Code should be limited to the transposition of AVMSD), to the extent that the Commission wishes VSPSs to address specific types of harms within these broad categories and/or provide additional context around particular types of harmful content that would fall within each of these categories, this may be more appropriately achieved by way of non-binding guidance, not by a binding code. Guidance could provide examples to VSPSs on what may be deemed to be the “main” harms to address and then inform the Commission’s enforcement priorities, without limiting the scope of the Code or primary legislation.</p> <p>This is, in our view, important given that different categories of harm are already addressed in broader regulatory frameworks which are specifically designed to address particular classes of content such as Regulation (EU) 2021/784 on addressing the dissemination of terrorist content online⁶ (the TCO Regulation) and the proposal of the European Parliament and of the Council laying down rules to prevent and combat child sexual abuse (the CSAM Proposal).⁷ It is also the case that the DSA itself, in respect of VLOPs and VLOSEs, expressly requires systemic risks arising from their services and systems to be assessed in a detailed fashion and then mitigated (Articles 34 and 35 DSA).</p>
<p>Question 2:</p> <p>What types of online harms do you think should attract the most stringent risk mitigation measures by VSPS? How could we evaluate the impact of different types of harms e.g. severity, speed at which harm may be caused? Is there a way of classifying harmful</p>	<p><u>What types of online harms do you think should attract the most stringent risk mitigation measures by VSPS?</u></p> <p>While TikTok takes all forms of online harms seriously and works continuously to ensure a safe community for its users, TikTok recognises that not all online harms are the same. We are of the opinion that the most stringent risk mitigation measures adopted by VSPSs should be in relation to illegal content. TikTok has zero tolerance for child sexual exploitation and other such illegal content. Our Community Guidelines apply to everyone and everything on our platform, and they often match, and sometimes go beyond, local law requirements. In addition, VSPSs have a degree of certainty</p>

⁶ Regulation (EU) 2021/784 of the European Parliament and of the Council of 29 April 2021 on addressing the dissemination of terrorist content online

⁷ Proposal for a Regulation of the European Parliament and of the Council laying down rules to prevent and combat child sexual abuse

content that you consider it would be useful for us to use?

Question 3:

Do you have reports, academic studies or other relevant independent research that would support your views? If you do, please share them with us with links to relevant reports, studies or research.

as to what constitutes illegal content and this online content can be clearly defined for the purposes of implementing strict mitigation measures by VSPSSs.

Our Community Guidelines make clear to our users that we remove this content and in certain circumstances, the below violations would result in an immediate account ban on our platform. We also recently introduced in the existing reporting function an additional [dedicated channel](#) for users to report illegal content to alert us to content they believe breaches the law.

- Post, promote, or facilitate youth exploitation or child sexual abuse material.
- Promote or threaten violence.
- Post or promote content that depicts non-consensual sex acts such as rape or molestation.
- Post content that facilitates human trafficking.
- Post content that depicts real-world torture.

Our content policies deem this type of content to be particularly egregious and, to the extent that such risk mitigation measures are aligned with DSA requirements, we would be of the opinion that these types of content would potentially warrant stringent risk mitigation measures in the Code. Such measures should align with, and not go beyond, the requirements of the DSA and platforms should be able to point to any relevant DSA compliance and/or risk mitigation measures to demonstrate their compliance with any such measure under the Code.

How could we evaluate the impact of different types of harms e.g. severity, speed at which harm may be caused? Is there a way of classifying harmful content that you consider it would be useful for us to use?

Again, and as per our response to Question 1, broad categories of harmful online content are included in both the AVMSD and OSMR due to the subjectivity associated with determining what may or may not amount to harmful online content, and the cases-by-case assessment that will need to be made by VSPSSs to combat such online harms. These broad categories already facilitate a degree of classification, while also achieving the balance of recognising the case-by-case assessments required by providers to ensure that they are appropriately regulating content online.

However, to the extent that the Commission wishes to further classify harmful content via the Code, an appropriate and balanced way to achieve this is to prioritise harms towards children in accordance with AVMSD standards/categories. Not only does this approach effectively implement the AVMSD standards and requirements on a national level, it also aligns with ongoing public commitments made by the Commission to prioritise the protection of children. For example, the Commission's work programme states that the Code will "...include measures that videosharing platforms must take to address matters such as the protection of minors from harmful video content, hate speech directed against groups with protected characteristics, and criminal offences, including those related to terrorism, child sex abuse material and racism"

	<p>(emphasis added). The Commission’s website also includes aims such as “<i>protecting children and all of us from harmful content</i>” and previous statements published by the Commission also highlight that Codes, in particular, will be used to “<i>protect children from age-inappropriate content</i>”.</p> <p>However, regard must also be had to other existing frameworks that address harm towards children; the Irish DPC’s fundamentals for a child-oriented approach to data processing (the Fundamentals); the UK age appropriate design Code (the UK Age Appropriate Code); the planned EU Code of Conduct for the Protection of Minors(the Code for the Protection of Minors); the European Commission’s new strategy for a better internet for kids(the EC’s Strategy for Kids). Again, any classification of harm must be aligned to these regimes.</p>
<p>Section 4: Overall Approach to the Code</p>	
<p>How prescriptive or flexible should the Code be?</p> <ul style="list-style-type: none"> ● Option 1 – A very detailed, prescriptive Code ● Option 2 – A very high-level Code ● Option 3 – A mixed approach 	<p>For the reasons outlined in more detail below in our response to question 4, namely that the Code should be sufficiently flexible in order to effectively regulate a variety of different VSPSs and that it should avoid conflicting with other EU regulatory regimes, we consider that Option 3 of introducing a very high-level Code, supplemented by non-binding guidance, is the most appropriate, proportionate and effective approach. This approach would involve setting out the categories of harm for VSPS providers to address and then obliging those providers to take appropriate measures to reduce the risk of harm in general terms, with the guidance providing the Commission’s view on the factors that can be taken into account when assessing compliance with the principles. Indeed, this is the approach that TikTok has operated under as part of the UK’s AVMSD implementation and we have found that it has worked well in practice.</p> <p>We would discourage the Commission from introducing a very detailed, prescriptive Code under Option 1 as this would run the risk of conflict arising between the Code and obligations VSPSs are already subject to under existing regulations. Additionally, as further described below, the more prescriptive the measures the more likely they are to be unworkable for certain VSPSs and therefore ineffective, as one size will not fit all. For these reasons, in our view the Commission should, at a minimum, adopt the approach at Option 2, but not go beyond the approach described by Option 3</p>
<p>Question 4:</p> <p>What approach do you think we should take to the level of detail in the Code? What role could non-binding</p>	<p><u>What approach do you think we should take to the level of detail in the Code?</u></p> <p>We understand from the Call for Inputs that the Commission intends to adopt one Code for all VSPS providers (at least initially). In light of the fact that VSPSs vary in terms of their users, size and the kinds of content they make available, in</p>

guidance play in supplementing the Code?

our view it will be extremely important that the Code takes a principles-based approach to regulation to ensure effective, practical, proportionate, and legally robust regulation.

We consider that the Code should set out at a high-level its objectives and principles and should oblige VSPSs to take appropriate measures to reduce the risk of harm in general terms. This would allow VSPS providers a degree of necessary flexibility to determine what measures they must implement to comply with the regime in the most effective manner. Such an approach would be consistent with the requirements of the AVMSD which provides that the measures VSPS are required to take should be “*practicable and proportionate, taking into account the size of the video-sharing platform service and the nature of the service that is provided*”. This approach is also fully in line with the DSA’s risk-based approach to the regulation of providers of online platforms.

If the Code is too prescriptive, there is a risk that it will not be proportionate and adaptable to the different types of VSPSs that will be regulated and, as a result, will be less effective and/or will risk cutting across the harmonised approach mandated by the DSA. Indeed, it is worth noting that Article 28b(3) of the AVMSD sets out a list of measures which VSPS providers should be required to take “*as appropriate*”. This expressly recognises that each of the measures listed in Article 28b(3) may not be appropriate for all types of VSPSs and that the measures VSPSs implement to achieve the objectives of the legislation may vary depending on the nature and scale of each service.

In addition, there is a clear risk that a prescriptive Code with detailed obligations on VSPSs may run the risk of discouraging or disincentivising VSPSs from implementing new protective innovations for online safety which would, in turn, undermine the goals to be achieved by the AVMSD. Further, by being too prescriptive there is an inherent risk that legislation cannot keep pace with the rate of change in innovation and the development of technology. Indeed, an overly prescriptive code would be a barrier to entry to the market and potentially impede the ability of new start-ups and smaller enterprises into the industry.

By taking a high-level approach to the Code, the risk of conflict with other EU regulatory regimes (such as the DSA) is greatly reduced. Whereas, a prescriptive approach risks overlapping/conflicting with other regimes, a principles-based approach would maximise the potential for synergies in how platforms comply with it and other EU regulatory regimes at the same time.

What role could non-binding guidance play in supplementing the Code?

We consider that non-binding guidance may play a role in supplementing the Code, where necessary and appropriate for the Commission to indicate its regulatory expectations on certain specific issues (indeed, this is the approach [Ofcom](#) adopted in the UK and which we have found useful). As noted above, in order to effectively regulate in a clear and transparent way a variety of different types of VSPSs, the Code will need to be at a high-level. However, in certain cases

	<p>non-binding guidance may further assist VSPs with understanding certain obligations under the Code and with specifying the matters which the Commission requires VSPs to prioritise (see response to question 1 above). Such an approach would allow the Commission to set out their expectations without the negative effect of an overly prescriptive Code which may have the unintended consequence of being technically impractical for varying platforms.</p> <p>By providing for these measures in non-binding guidance this would allow VSPs to comply with these provisions in a manner that is tailored to specific risks presented by their platform. This would also be consistent with Recital 49 of the revised AVMSD which provides that it is appropriate to involve VSPs as much as possible when implementing the appropriate measures to be taken pursuant to the AVMSD and that co-regulation should therefore be encouraged.</p> <p>This guidance could adopt a similar approach to the various guides produced by Ofcom, the UK's communications regulator, to contextualise the VSP regime and which are helpful in understanding how best to go about complying with the high-level principles.</p>
<p>Question 5:</p> <p>What do you think would be the most effective structure for the Code? What are the most important factors we should consider when we decide how to structure the Code?</p>	<p><i>What do you think would be the most effective structure for the Code?</i></p> <p>As noted above, we consider that the Code should be high-level. The most effective structure for such a Code would, in our view, be a thematic structure based on how the different elements relevant to VSPs would be impacted by the appropriate measures set out in Article 28b(3) of the AVMSD. We also believe it may be useful for the Code to set out any guiding and overarching principles at the beginning. This would align with the high-level example outlined in the Call for Inputs.</p> <p><i>What are the most important factors we should consider when we decide how to structure the Code?</i></p> <p>In our view, one of the most important factors that should be considered when designing the structure of the Code is adaptability. In order for the Code to be effective, it is crucial that the Code is structured in such a manner that its requirements are adaptable to all of the different types of VSPs that it will regulate. Additionally it will be important that the Code is structured in an easily adaptable manner so that it will stand the test of time and effectively regulate VSPs in the future. Again, a high-level structure as suggested in the Call for Inputs would most likely achieve these objectives.</p>
<p>Question 6:</p> <p>How should we design the Code to minimise the potential for conflict and</p>	<p>The DSA expressly cautions Member States against adopting additional national laws on the matters covered by the DSA. The DSA recognises requirements addressing online safety and the dissemination of illegal content online as an</p>

<p>maximise the potential for synergies in how platforms comply with it and the DSA?</p>	<p>area which should be fully harmonised under the DSA and accordingly provides that Member States should not adopt national measures dealing with this area.⁸</p> <p>To ensure legally robust measures and the effective implementation of the AVMSD, it is crucial that the Code does not conflict in any manner with the DSA. As the Commission is aware, VSPSs have to comply with obligations for online platforms under the DSA and some, like TikTok, will be subject to additional obligations as VLOPs under the DSA. We suggest that this should be explicitly acknowledged in the Code by way of a general principle clarifying that where any obligation contained in the Code overlaps to any degree with obligations contained in the DSA, a failure to comply with that obligation will not give rise to a contravention provided that a provider is in compliance with its DSA obligations and that the DSA should take precedence in the event of any conflict between DSA and the Code.</p> <p>With regard to the suggestion in the Call for Inputs that the Code might “...impose additional and/or more detailed requirements on VSPS providers” than the requirements under the DSA, we would note that the Recitals to the DSA make clear that no additional national requirements should be introduced relating to matters falling within its scope. This would cut across the harmonised approach at an EU level sought to be achieved by the DSA. Accordingly, if the Code were to impose additional and/or more detailed requirements on VSPS providers, this would require careful consideration in order to ensure that the matters covered by the Code do not relate (with a high degree of certainty) to the matters that fall within the scope of the DSA.</p> <p>We very much agree with the Call for Inputs that it would be helpful to design the Code in a way that maximises the potential for “synergies” in how platforms comply with it and the DSA. In order to do so, it will be important that the Code not only avoids conflict with the DSA, but that it also takes into account the obligations that VSPS providers are already subject to under existing regulations, so as to ensure they are not subject to duplicative regulatory requirements which would seriously undermine the objectives sought to be achieved by the OSMR.</p> <p>Further harmonisation with other EU content regulation will be of key importance and it will be important that the Code ensures effective harmonisation with other regulatory regimes such as the TCO Regulation, the CSAM Proposal, the Strengthened Code of Practice on Disinformation 2022 (the COPD) and any new codes that may be introduced under the DSA. Where VSPSs are already subject to existing obligations under other regulatory regimes and these obligations assist VSPSs in complying with the objectives of the Code, this should be expressly acknowledged.</p>
<p>Question 7:</p>	<p>In our view, the Code should address video content only.</p>

⁸ Recital 9 DSA

<p>To what extent, if at all, should the Code require VSPS providers to take measures to address content connected to video content?</p>	<p>The AVMSD requires that VSPS providers take appropriate measures in respect of audiovisual content. It does not require VSPSs to take appropriate measures in relation to content connected to audiovisual content and therefore any regulation of connected content would go beyond the scope of the AVMSD. In addition, this approach, if adopted in Ireland, would align with the transposition of the AVMSD in the majority of EU Member States, being that national transposing measures very much correspond to the provisions of the revised AVMSD itself. In particular, it does not appear that there has been significant further elaboration on, or introduction of, stricter obligations for VSPSs⁹. To ensure legal clarity and to avoid potential patchwork implementation across the EU, the Code should avoid deviating from the requirements of the AVMSD.</p> <p>As such, we suggest that the primary focus of the Code is limited to addressing audiovisual content only, which aligns with the majority of Member States in the EU that have successfully transposed the AVMSD.</p>
<p>Section 5: Measures to be taken by Video-Sharing Platforms</p>	
<p>Preliminary TikTok comments on section 5</p>	<p>Before addressing each of the specific questions in section 5, we want to take this opportunity to reiterate the reasonable and proportionate approach that needs to be taken when considering the Article 28b(3) measures. In particular, it is important to at all times bear in mind the ‘appropriateness’ element of Article 28b(3) of the AVMSD. The implementation of ‘appropriate’ measures expressly recognises that each of the measures listed in Article 28b(3) may not be relevant or appropriate for all types of VSPSs and that the measures VSPSs implement to achieve the objectives of the legislation may vary depending on the nature, extent and scale of each service.</p>
<p>Question 8:</p> <p>How should CnaM ask VSPS providers to introduce a feature that allows users to declare when videos contain advertising or other type of commercial communications? Should the Code include specific requirements about the form in which the declaration should take? What</p>	<p><i>How should CnaM ask VSPS providers to introduce a feature that allows users to declare when videos contain advertising or other type of commercial communications?</i></p> <p>TikTok recognises the importance of ensuring that advertisements and other types of commercial communication are appropriately disclosed and recognised as such by users.</p> <p>Taking a principles-led approach, we do not consider that the Code should go beyond setting out that such a feature is required. This would then allow platforms to comply in a manner that is most effective for their individual service and user base. It would also allow platforms to demonstrate compliance by pointing to existing features and tools.</p>

⁹ EAO Publication

current examples are there that you regard as best practice?

In implementing measures to address this issue, we note for context that this is also an area which is regulated by the DSA (e.g. in Article 26) which requires service providers to ensure that users of online platforms can (i) identify advertisements, and related information including the identify of the advertiser, in a “clear, concise and unambiguous manner”; and (ii) declare other types of commercial communications, allowing them to be identified to other users in a “clear and unambiguous” way. It will be important that any requirements introduced by means of the Code align with the DSA obligations.

TikTok already has a number of tools in place to ensure that advertisements and other commercial communications are clearly and effectively labelled for users. TikTok is also transparent with users about its approach and we have included links to the simple, concise guides we make available to users:

- All advertisements on the TikTok platform are arranged and delivered through TikTok’s Ads Manager and other bespoke tools for advertisers. A prominent “Ad” label (or local equivalent) is automatically applied when the advertisement is displayed on the TikTok platform.
- TikTok also provides creators with accessible and user-friendly means to identify other types of commercial communication. In particular, when posting content that promotes a brand, product, or service on TikTok, the creator is required to turn on the content disclosure setting. When the creator uses these tools to identify their content, TikTok will automatically label the content using appropriate labels (for instance, “paid partnership” or “promotional content”).

These labels reflect standard industry practice and would be similar to disclosures used by other intermediary services that facilitate the posting of commercial communications through video content.

We agree with the Commission that commercial communications are an important source of funding for content creators and the Code should not prohibit or inhibit legitimate forms of commercial communications. We believe that our existing promotions features and guidelines achieve the balance which the Commission seeks to achieve i.e. respecting a content creator’s source of income, while clearly letting users know when they are being targeted with commercial messages.

Should the Code include specific requirements about the form in which the declaration should take? What current examples are there that you regard as best practice?

As detailed above, we believe general principles are most appropriate. A greater level of specificity is unlikely to be platform neutral and runs the risk of cutting across the DSA’s provisions on this issue. We also note that the European

	<p>Commission is required to encourage and facilitate specific standards for the required markings under Article 44, and codes of conduct for online advertising under Article 46 of the DSA.</p> <p>Additionally, as is acknowledged in the EAO Publication, the requirements of the AVMSD relating to the declaration of advertising/commercial communications have generally been transposed verbatim in other jurisdictions. We support this approach and would encourage the Commission to implement the AVMSD in a manner consistent with the majority of other EU Member States.</p>
<p>Question 9:</p> <p>How should we ask VSPS providers to introduce and design a flagging mechanism in the Code? How can we ensure that VSPS providers introduce the mechanism in a user-friendly and transparent way? How should we ask VSP Providers to report the decisions they've made on content after it has been flagged? To what extent should we align the Code with similar provisions on flagging in the DSA?</p>	<p>How should we ask VSPS providers to introduce and design a flagging mechanism in the Code?</p> <p>Our recommendation would be to focus on high-level principles, supported by non-binding guidance that would assist platforms in understanding the expectations of the Commission when it comes to reporting mechanisms. For example, the Commission could set the overall principle that reporting and flagging mechanisms should be 'user-friendly' and transparent and then provide separate guidance on what would inform its assessment of whether such mechanisms met the principles.</p> <p>To what extent the Code should be aligned with the provisions on reporting under the DSA, we would reiterate that the Code should not conflict with the DSA on any aspect and in particular, any attempt to be more prescriptive than the DSA provisions would cause confusion around VSPS obligations.</p> <p><i>Overview of TikTok's approach</i></p> <p>TikTok is focused on providing a safe experience for all of its users and works to maintain a safety by design approach to protect users' safety and well-being. As part of this approach, TikTok has always had in place functionalities that enable users to flag any content on its service that users consider might violate TikTok's Community Guidelines. If a member of our community sees anything that they believe violates these guidelines, they can report content or a profile directly from the app or desktop computer in a user-friendly manner. We then assess and take action against content reported to us. We consider this an important mechanism for user empowerment which further facilitates us in creating a safer environment on our platform.</p> <p>As part of requirements under the DSA, we provide our community with an additional option, 'Report Illegal Content', in the existing reporting function to alert us to content they believe breaches the law.</p>

How can we ensure that VSPS providers introduce the mechanism in a user-friendly and transparent way?

The Commission will be aware that the DSA places an obligation on intermediary services to provide online interfaces which should be user-friendly and easily accessible. This topic may benefit from non-binding guidance by the Commission akin to that set out in Ofcom’s [‘Video Sharing Platform Guidance.’](#)

We provide users with [simple, intuitive ways](#) to report/flag content in-app for any potential violation in any of the official languages of the European Union.

- By ‘long-pressing’ (e.g., clicking for 3 seconds) on the video content and selecting the “Report” option.
- By selecting the “Share” button available on the right-hand side of the video content and then selecting the “Report” option.

The user is then shown categories of reporting reasons from which to select (which align with the harms our CGs seek to address). We have also recently implemented an additional option to enable users to report suspected illegal content in line with our requirements under DSA.

How should we ask VSP Providers to report the decisions they’ve made on content after it has been flagged?

We note that Article 17 of the DSA requires hosting services to provide clear and specific statements of reasons to users affected by certain actions taken (such as the removal of content) based on content moderation decisions. Additionally, Article 20 DSA requires online platforms to provide recipients of the service, for a period of at least six months following the decision, with respect to flagged content, the access to an effective, user-friendly internal complaint-handling system, which enables the complaints to be lodged electronically and free of charge.

To assist the Commission in determining what may constitute best practice for reporting decisions on content after it has been flagged, we have provided information on TikTok’s existing approach which supports reporting of content made both under our community guidelines and illegal content under DSA.

In keeping with our commitment to ensuring procedural fairness, we seek to provide notifications to community members if they have violated our rules. If a user posts content that we do not allow or is determined to be illegal, they will be notified in the app along with the violation reason. If a user’s account has been banned because of a violation, they will receive a banner notification when they next open the app, informing them of this account change. If a user receives a notification of a content violation or account ban and believes that it was done in error, then they can appeal the decision.

	<p>Users can view the status of their appeal in the in-app Safety Center, as well as the status of any reports they have filed about other content or accounts.</p>
<p>Question 10:</p> <p>What requirements should the Code include about age verification and age assurance?</p> <ul style="list-style-type: none"> • What sort of content should be shown by default to users who are logged out or in private browsing mode and whose age cannot be verified or assured? • What evidence is there about the effectiveness of age estimation techniques? What current practices do you regard as best practice? • Where accounts are not age verified should default privacy settings be used, should content default to universal content and should contact by others be more limited? 	<p>We are deeply committed to the safety and privacy of our users, especially our younger users. In particular, we are committed to preventing under 13's accessing our platform and to continuing to enforce our policy on the platform by detecting and removing younger users who are not old enough to access the platform.</p> <p>Preventing underage people from creating an account</p> <ul style="list-style-type: none"> • TikTok has a 12+ rating in the App Store and Google Play, which enables parents to use device-level controls to block their teens from downloading TikTok. • To help keep people from using our platform if they're not yet old enough to do so, we've designed a neutral, industry-standard age gate that requires people to fill in their complete date of birth. • If someone tries to create an account but does not meet our minimum age requirement, we suspend their ability to attempt to create another account using a different date of birth. <p>Removing suspected underage accounts</p> <ul style="list-style-type: none"> • Our commitment to enforcing our minimum age requirements does not end at the age gate, and we take a number of additional approaches to identify and remove suspected underage account holders. • We train our safety moderation team to be alert to signs that an account may belong to someone under the age of 13. We also use other information provided by our users, such as keywords and in-app reports from our community, to help surface potential underage accounts. • When our safety team believes that an account may belong to an underage person, the account will be suspended. • If an account is being reviewed by one of our moderators for another violation and the moderator identifies that the account holder appears to be under 13, the account will be removed or flagged for further review by our underage moderation team. • To bring more visibility to the actions we take to protect minors, we are the only major platform to regularly disclose the number of accounts we remove from the full TikTok experience for potentially belonging to an underage person. <p>TikTok is committed to exploring innovative solutions in these areas. We believe the industry should work toward accessible, robust, privacy preserving options. Currently, there is no "silver bullet" age assurance solution that can be</p>

	<p>rolled out across all platforms in a way that fully accounts for a younger user's right to privacy. If such a solution was available, we would be keen to explore it further. In the meantime, we are working towards further enhancing our age assurance strategy. We recognise that this is a dynamic issue, and solutions may evolve across operating systems and platforms. Practical examples of age assurance measures that the Commission views as effective and privacy protective would also be welcomed as part of any guidance documents to be published.</p> <p>We would note that any age verification / age assurance requirements introduced via the Code should also be aligned with, and take account of, other existing frameworks which address these matters, including the Irish DPC's Fundamentals; the UK Age Appropriate Code; the Code for the Protection of Minors; the EC's Strategy for Kids.</p>
<p>Question 11:</p> <p>What requirements should the Code have in relation to content rating?</p> <ul style="list-style-type: none"> • What do you consider to be current best practice? • What experiences have you had using content rating systems on platforms and do you think they have been effective? • What steps could we ask VSPS to take to ensure content is rated accurately by users? 	<p>The area of classifying content, or ensuring that that content is viewed by appropriate audiences, is highly complex. This further underscores the necessity for a principles-based approach in this area rather than prescriptive solutions. An approach by or for one intermediary service may not work on another. As such, TikTok's view is that the focus should be on ensuring that such services are empowered to demonstrate compliance in the way that is most effective for their user base.</p> <p>In case helpful, we have set out below an overview of the approach TikTok has taken and considers to be effective in our context.</p> <p>TikTok has developed Content Levels which organises content based on thematic maturity and giving users choice based on their personal preferences. It is important to emphasise that this, and any, content on the platform must comply with our Community Guidelines. Within these strict policies however, we understand that people may want to avoid certain categories of content based on their personal preferences - for example, fictional scenes that may be too intense. Or, for our teenage community members, some content may contain mature or complex themes that may reflect personal experiences or real-world events that are intended for older audiences. Our approach is similar to what we see in the film, television broadcast, and gaming industries. We are drawing closely on the kinds of standards already in use around the world. We have focused on further safeguarding the teen experience first and we plan to add new functionality to provide detailed content filtering options for our entire community so they can enjoy more of what they love.</p>
<p>Question 12:</p> <p>What requirements should the Code have in relation to parental control</p>	<p><i>What requirements should the Code have in relation to parental control features?</i></p> <p>The Commission should consider adopting an approach to parental control consistent with a substantial number of other EU Member States and be consistent with obligations of data protection, a teenager's right to autonomy and the freedom</p>

features? How can we ensure that VSPS providers introduce the mechanism in a user-friendly and transparent way? Can you point to any existing example of best practice in this area? Should parental controls be 'turned-on' by default for accounts of minors or where age is not verified?

of expression. As is stated in the EOA Publication, most national legislation transposes verbatim the provisions of the AVMSD regarding measures to address content which may impair the physical, mental or moral development of minors.

How can we ensure that VSPS providers introduce the mechanism in a user-friendly and transparent way?

We invest in tools and resources to help parents, guardians, and families support their teens online. We have developed [settings](#) that can be enabled to manage a family's TikTok experience, including tools for filtering comments, blocking accounts, setting screen time limits, and disabling video downloads.

Can you point to any existing example of best practice in this area?

We welcome the opportunity to provide the Commission with information on our approach to this area which includes the parental tools TikTok has in place to protect users who are under the age of 18.

TikTok's 'Family Pairing' features let parents link their TikTok account to their teen's to enable a variety of content, privacy, and well-being settings. We encourage caregivers to discuss the Family Pairing features with their teens and to collaborate in identifying the most appropriate content experience for the teen in question.

Even without Family Pairing enabled, parents can help their teens enable our app's Screen Time offerings, including Daily Screen Time and Restricted Mode. Our screen time management tools seek to strike a balance between autonomy, expression, and broader digital well-being with additional protections in place for teens which we consider to be reasonable and proportionate in regard to their relative age and maturity.

Family Pairing on TikTok allows parents and young users to customise their safety settings based on individual needs. A parent or guardian can link their TikTok account to their teen's account and collaborate with their teen on various empowerment tools including:

Daily Screen Time

- A screen time management setting that allows you to manage your app usage. It lets you set a daily screen time limit so that you get notified when you reach that time on TikTok. You can turn this setting on and off at any time. If you're between the ages of 13 and 17, the setting is turned on by default to 1 hour

Filter video keywords

- Selecting keywords or hashtags to exclude from a child's For You and Following feeds, as well as manage the visibility of keywords list.

Restricted Mode

- Restricted Mode on TikTok limits exposure to content that may not be suitable for everyone, for example, because it contains mature or complex themes. Some features will be unavailable under Restricted Mode, such as the Following feed and gifting on LIVE.

Linked account activity

- Parents can receive notifications about their teen's activity, such as if their accounts get unlinked, by turning on customised updates and more push notifications.

Search

- Enhanced controls over search for videos, hashtags, or LIVE videos on TikTok.

Discoverability

- Control over whether the account is private or public.

Suggest account to others

- Control over whether the account can be recommended to others.

Direct Messages

- Control over who can send messages to their teen, or turn off direct messaging completely. Direct messaging on TikTok is available only to registered account holders aged 16 and older. Direct messaging is automatically turned off for registered accounts between the ages of 13 and 15.

Liked videos

- Control over who can view liked videos.

Comments

- Control over who can comment on videos.

Screen time dashboard

	<ul style="list-style-type: none"> • In family pairing; provides summaries of time on the app, the number of times TikTok was opened, and a breakdown of total time spent during the day and night. <p>Mute notifications</p> <ul style="list-style-type: none"> • Enables parents to set a schedule to mute notifications for their teen. Accounts aged 13-15 already do not receive push notifications from 9pm and accounts aged 16-17 have push notifications disabled from 10pm. <p><i>Should parental controls be ‘turned-on’ by default for accounts of minors or where age is not verified?</i></p> <p>See our response to question 10. In addition, any 'default on' parental settings as described by the Commission must comply with data protection rules, and in particular with the principles of proportionality, transparency and security of the affected data subjects. The Commission may also want to explore whether and to what extent 'default on' parental controls would comply with the GDPR. There is also an inherent technical challenge in legally identifying a parent/guardian.</p>
<p>Question 13:</p> <p>What requirements should the Code contain to ensure that VSPS provide for effective media literacy measures and tools?</p>	<p>Overall, TikTok's view is that a principles based approach to media literacy would involve setting out the broad requirements the Commission would expect when it comes to media literacy and, if necessary, that it supplements the principles with guidance. We consider this will be the most effective approach as different platforms will necessitate differing approaches depending on their scale, user base, content types etc. In case helpful, we have set out below an overview and some examples of the approach TikTok takes in this area.</p> <p>At TikTok, we take a broad view of media and digital literacy, and adopt a range of measures to enhance media literacy and generate awareness of risks and safety issues for our users. We place a considerable emphasis on generating awareness and helping to foster the development of skills for users to critically assess and understand information in an online context. This approach is not limited to users of the platform, but also includes media literacy resources for educators, parents and caregivers so that they are better placed to safely and responsibly navigate their online experiences in connection with the platform.</p> <p>In order to raise awareness among our users of specific topics and empower them, we run a variety of on and off-platform media literacy campaigns. Our approach may differ depending on the topic. We localise certain campaigns (e.g., for elections) in that we collaborate with national partners and use language that the local audience can best connect with.</p>

Resources

- [Safety Centre](#): Within our Safety Centre, we have several guides on a range of topics, including a page on Digital Well-Being, which discusses media literacy and encourages users to question the source of the information they consume.
- [Help Centre](#): Our Help Centre provides users with accessible 'how to' explanations of our user experience to allow them to learn about the Platform and troubleshoot issues. The “Safety” page in particular contains the following subsections:
 - “Account and user safety”: which further explains “Content violations and bans”; our “Community Guidelines”; Account Safety; and “user Safety”; and others; and
 - “Report a problem”: which further explains how users (and non-users) can report content to TikTok (as well as reporting suspected underage users).
- [Transparency Reporting](#): Within the Our Commitments page, we have user-friendly articles explaining TikTok’s approach to Keeping People Safe, which includes separate articles that explains *Our approach to content moderation*.
- [Online Newsroom](#): We use our Newsroom posts to communicate with our community transparently and to build and maintain trust. We publish a range of posts in our Newsroom in which we seek, among other things, to generate awareness on safety and content related issues.

Campaigns

Since 2020, on topics such as [Covid-19](#), [Covid-19 Vaccine](#), [Holocaust Denial](#), MonkeyPox and War in Ukraine, we deployed a combination of a number of in-app intervention tools such as video notice tags, search interventions and public service announcements. In the last 6 months alone, we have developed, together with our fact-checking partners, and rolled-out eight localised media literacy campaigns on the war in Poland, Slovakia, Romania, Ukraine, Hungary, Estonia, Latvia and Lithuania. Users searching for keywords relating to the war are directed to tips, prepared in partnership with our fact checking partners, to help users identify misinformation and prevent the spread of it on the platform. We have also launched a [climate change search intervention tool](#), which redirects users seeking out climate change-related content to authoritative information (i.e. UN [resources](#)) and encourages them to report any potential misinformation content they encounter.

Question 14:

How should we ask VSPS providers to address online harms in their terms and conditions in the Code, including the harms addressed under Article 28b? How should key aspects of terms and conditions be brought to users' attention? What examples are there of best practice in relation to terms and conditions, including content moderation policies and guidelines?

How should we ask VSPS providers to address online harms in their terms and conditions in the Code, including the harms addressed under Article 28b?

The EAO Publication identified that, as regards Article 28b(3)(a) of the AVMSD requiring VSPSs to take appropriate measures relative to the inclusion and application of their terms and conditions, the majority of Member States have transposed this provision “*by citing the provisions of the AVMSD verbatim*”. The EAO Publication clarifies that, in doing this, the emphasis is put on the easiness, understandability and simplicity, as well as the accessibility, of VSPSs terms and conditions.

In our view that approach, as is consistent with the majority of other EU Member States, is the correct approach. Adopting a different approach risks patchwork implementation of the AVSMD across different EU jurisdictions and undermines the harmonised approach required by DSA.

What examples are there of best practice in relation to terms and conditions, including content moderation policies and guidelines?

We consider that our [Community Guidelines](#), our Terms of Services and other related documents (together, our **T&Cs**) reflect best practice and would encourage the Commission to take these into account in the drafting of the Code. TikTok has worked hard to ensure that our Community Guidelines bring issues relating to harmful online content to users' attention in a transparent and easily accessible manner. They are informed by international legal frameworks, industry best practices, and input from our community, safety and public health experts, and our regional Advisory Councils. To assist the Commission in this regard, we have sought to outline below the elements of our T&Cs as representing aspects of the best practice in the industry and which we consider are relevant to any requirements under the Code:

T&Cs and related documents should be easy to navigate and user-friendly

Our T&Cs are structured in such a way to allow any individual to be able to easily navigate and find the relevant information that they are looking for. We use clear, simple and concise wording and we make our T&Cs available in 25 European languages. We have also produced a [summary](#) of our T&Cs.

Our Community Guidelines are organised by topic area, with each rule in **bold**. We first explain in brief what we don't allow, and we then provide more details, such as definitions and the range of actions we might take. Under each section a user can click for more information where a user can find definitions, specific examples, and clarifications to common questions about what is allowed. Our Terms of Service also includes a succinct and accessible “in short” section at the end of each provision, summarising the main points for users. We are always improving and evolving our policies and in

our [recent refresh](#) of our Community Guidelines, we made several enhancements including more detail about how we use informational labels, warnings, and opt-in screens.

Online harms should be clearly addressed in T&Cs and related documents

Our Community Guidelines have a navigation pane which is clearly positioned on the left-hand side of the webpage which lists the categories of online harm (as detailed under Article 28b(1)(a)-(c)), as well as other key areas such as our community principles and enforcement. A user can easily click on the relevant category and read the information outlined. These categories also align with the reporting reasons the user can select from when reporting content. A broad overview of the categories included in our [Community Guidelines](#) is as follows:

- [Mental and Behavioural Health](#) (Article 28b(1)(a) AVMSD) - this section includes information on suicide and self-harm, eating disorders and body image and dangerous activities and challenges (i.e. activities, trends or challenges that may lead to significant physical harm);
- [Youth Safety and Well-Being](#) (Article 28b(1)(a)) - this section contains information regarding TikTok's policies on content that may put young people at risk of exploitation, or psychological, physical, or developmental harm. This includes child sexual abuse material, youth abuse, bullying, dangerous activities and challenges, exposure to overtly mature themes, and consumption of alcohol, tobacco, drugs, or regulated substances;
- [Safety and Civility](#) (Article 28b(1)(b) & (c)) - this section contains information relating to the following categories of harmful and illegal online content:
 - Violent Behaviours and Criminal Activities
 - Hate Speech and Hateful Behaviours
 - Violent and Hateful Organisations and Individuals
 - Youth Exploitation and Abuse
 - Sexual Exploitation and Gender-Based Violence
 - Human Exploitation
 - Harassment and Bullying
- [Sensitive and Mature Themes](#) (Article 28b(1)(a)) - this section includes information on the following categories of harmful content: Sexual Activity and Services, Nudity and Body Exposure, Sexually Suggestive Content, Shocking and Graphic Content, Animal Abuse.

Permitted and prohibited use of the service (and consequences) should be clearly outlined in T&Cs

We believe TikTok's current approach of clearly but succinctly outlining in our [Terms of Service](#) what users can do (section 4.4) and what users cannot do (section 4.5) on the platform remains appropriate and reflects best practice. Under the

	<p>section of “<i>What users cannot do on the platform</i>”, the first point included is that users must not use the platform to do anything illegal including posting illegal content. This section also links directly to our Community Guidelines and states that they “<i>apply to everyone and to all content on the Platform</i>”. In our Community Guidelines, under each category listed above, what is “not allowed” and “allowed” are clearly outlined in each instance and specific examples are provided. Both the Terms of Service (Section 4.6) and the Community Guidelines outline to users that we have the right to remove or restrict access to any content if TikTok reasonably believes it is in breach of the Terms of Service or the Community Guidelines.</p>
<p>Question 15:</p> <p>How should we ask VSPS providers to address content moderation in the Code?</p> <ul style="list-style-type: none"> • Are there any current practices which you consider to be best practice? • How should we address automated content detection and moderation in the Code? 	<p>As noted in our responses above, we understand that one of the key objectives of the Code is the transposition of Article 28(b) of AVMSD and we believe this should be the Commission’s primary focus for the Code. In circumstances where the measures listed at Article 28b(3) of the AVMSD do not require the introduction of measures regarding content moderation, we consider this to deviate from the provisions of the AVMSD. As such, we do not believe the Commission should address content moderation in the Code. As noted in our response to Question 1, given that the transposition of the AVMSD is an urgent priority for Ireland, the Code should be limited to transposing Article 28b of the AVMSD, including to avoid any further delays to its transposition.</p> <p>We would also note that the DSA has introduced important content moderation requirements primarily aimed at transparency. Platforms have been required to include in their terms and conditions information on any restrictions that they impose on the use of their service. The Commission should therefore take the content moderation obligations under the DSA into account and should be particularly cautious about introducing any additional requirements in respect of content moderation on VSPSs, as this risks cutting across the matters regulated by the DSA which it seeks to harmonise at an EU level.</p>
<p>Question 16:</p> <p>What requirements should the Code include about procedures for complaint-handling and resolution, including out-of-court redress or alternative-dispute resolution processes?</p> <ol style="list-style-type: none"> 1. To what extent should these requirements align with 	<p><i>What requirements should the Code include about procedures for complaint-handling and resolution, including out-of-court redress or alternative-dispute resolution processes?</i></p> <p>As a preliminary point, we note that, as regards the manner in which the complaints handling requirements of the AVMSD have been transposed in other jurisdictions, the EAO Publication notes that the measures adopted in other jurisdictions predominantly reflect the requirements stipulated by the AVMSD, without the need to further strengthen or augment the obligation. These implementations more closely align with the requirements of Article 29b(3)(i) of the AVMS and the process suggested by this question, focusing on complaints regarding <i>how</i> platforms implement the measures required under Article 29b(3)(d) to (h). Under this approach, the focus of the complaints requirement is at a higher/structural-level, in particular on the manner in which a VSPS has complied with / implemented the measures rather than at a tactical level</p>

<p>similar requirements in the DSA?</p> <p>2. What current practices could be regarded as best practice?</p> <p>3. How frequently should VSPS providers be obliged to report to the Commission on their complaint handling systems and what should those reports contain?</p> <p>4. Should there be a maximum time-period for VSPS providers to handle user complaints and if so, what should that period be?</p>	<p>(e.g. has the VSPS implemented reporting and flagging measures in an effective way, rather than focusing on complaints regarding specific moderation decisions). From TikTok’s perspective, we agree that there is no need to further strengthen or augment the obligations provided for in the AVMSD and that the Commission should take a similar approach. In case the Commission elects to proceed in a broader manner, however, TikTok addresses the specific queries below.</p> <p>1. <i>To what extent should these requirements align with similar requirements in the DSA? What current practices could be regarded as best practice?</i></p> <p>As per our preliminary comments, VSPSs are subject to other related regulatory regimes in the EU which impose complaint handling and out-of-court settlement requirements - most particularly the DSA. As such, parallel complaints could be made in respect of the same harmful content, available on the VSPS, through any specific complaints mechanism required by the Code and the complaint mechanism required by the DSA. This means that there is a real risk of material overlap/conflict between the Code and the DSA if the Commission chooses to provide for similar complaint handling and out-of-court redress mechanisms in the Code. Of course, to the extent the Commission goes further than the strict requirements of Article 29b(3)(i), this increases the risk of DSA overlap/conflict.</p> <p>In order to avoid this real potential overlap and any confusion that would arise with users due to differing complaint mechanisms and redress procedures, we think it most appropriate that the Code include the greatest possible level of alignment with the DSA regime.</p> <p>Ultimately, the primary aim of these procedures is to ensure that they are user-centric i.e. affected or aggrieved users are able to effectively exercise their rights and avail of a simple, meaningful and timely redress mechanism when they have been negatively affected by harmful online content. We believe that the above suggested approach ensures that the user remains paramount with regard to a complaints/redress procedure.</p> <p>2. <i>How frequently should VSPS providers be obliged to report to the Commission on their complaint handling systems and what should those reports contain?</i></p> <p>The DSA requires providers of VLOPs to make publicly available on a six monthly basis a comprehensive report on content moderation and related practices and the underlying metrics (Article 15, Article 24 and Article 42). One of the required elements of such regular reporting requires TikTok to disclose the number of complaints it receives from users, the basis for them and the decisions taken on complaints, the median time needed for taking those decisions and the number of instances where those decisions were reversed.</p>
---	---

	<p>Additionally, TikTok regularly publishes comprehensive voluntary Transparency Reports to provide visibility into how we uphold our Community Guidelines and respond to law enforcement requests for information, government requests for content removals, and intellectual property removal requests.</p> <p>3. <i>Should there be a maximum time-period for VSPS providers to handle user complaints and if so, what should that period be?</i></p> <p>TikTok recommends that providers of intermediary services should be required to handle complaints in an efficient and timely manner in line with obligations under the DSA. Each complaint should be assessed on a case-by-case basis by the service. A prescriptive application of a time limit may reduce the efficacy of complaint handling as it may serve to disincentive VSPSs from properly considering the more complex issues that could be raised.</p> <p>As noted in our response above, the DSA imposes significant transparency obligations on intermediary services which includes information on complaints handling and as such we consider that these obligations, along with the complaints handling requirements of the DSA more generally, will ensure that complaints are handled in a timely and efficient manner. In these circumstances, the Commission should refrain from introducing any more prescriptive timeframes for the handling of complaints.</p>
<p>Question 17:</p> <p>What approach do you think the Code should take to ensuring that the safety measures we ask VSPS providers to take are accessible to people with disabilities?</p>	<p>In line with our comments above in respect of a general user and service risk-based approach, we do not believe the Code ought to deal in a prescriptive manner with this particular issue. In this regard, we also note that the European Commission is required to encourage and facilitate codes of conduct for accessibility under Article 47 of the DSA which many VSPSs will also need to take account of in carrying out DSA risk assessments.</p>
<p>Question 18:</p> <p>What approach do you think the Code should take to risk assessments and safety by design? Are there any examples you can point us towards</p>	<p>The DSA specifically requires that providers of VLOPs and VLOSEs carry out risk assessments to identify, analyse and assess systemic risks arising from the design, functioning or use made of their services, including the risk of dissemination of illegal and harmful content through their services (Article 34 DSA).</p> <p>As the Commission will be aware, the DSA risk assessments have to consider systemic risks including, e.g, the dissemination of illegal content and negative effects on fundamental rights. In particular, VLOPs and VLOSEs are required to take into account the following factors as part of their risk assessments in determining whether they influence</p>

<p>which you consider to be best practice?</p>	<p>systemic risks stemming from their platform and services:</p> <ul style="list-style-type: none"> (a) the design of their recommender systems and any other relevant algorithmic system; (b) their content moderation systems; (c) the applicable terms and conditions and their enforcement; (d) systems for selecting and presenting advertisements; and (e) data related practices of the provider. <p>To the extent that the above factors do influence systemic risks on their services, the VLOP/VLOSE will have to ensure that appropriate mitigation measures are implemented. Such risk assessments must be submitted to the Digital Services Coordinator of Establishment (i.e. the Commission) and the European Commission without undue delay upon completion (Article 42(4) DSA).</p> <p>As we have noted above, the DSA expressly provides that Member States should not adopt national measures in any of the areas regulated by the DSA to avoid fragmentation of the internal market and to ensure legal certainty.¹⁰ The DSA only places an obligation to carry out risk assessment on VLOPs and VLOSEs and not on other providers of intermediary services. This reflects the graduated approach of the DSA. Any national measures that attempt to undermine this approach would also cut across the harmonised and graduated approach sought to be achieved by the DSA.</p> <p>In light of the above, if any form of risk assessments are to be dealt with under the framework of the OSMR, these should be provided for in non-binding guidance and should build upon DSA risk assessment obligations solely to the extent required to meet the aims of the AVMSD and applicable only to those VSPSs that are not also VLOPs or VLOSEs under the DSA.</p>
<p>Question 19:</p> <p>How do you think that cooperation with other regulators and bodies can help us to implement the Code for VSPS?</p>	<p>Given the overlap in regulatory regimes, there may be benefit in the Commission being cognisant of the approach taken in various areas by other regulators (e.g. the European Commission in respect of the DSA, the Commission for Communications Regulation in respect of electronic communications services and the Data Protection Commission in respect of data privacy and data subject rights).</p>

¹⁰ see Recital 4 and 9 DSA.

Question 20:

What approach do you think we should take in the Code to address feeds which cause harm because of the aggregate impact of the content they provide access to?

- Are there current practices which you consider to be best practice in this regard?

What approach do you think we should take in the Code to address feeds which cause harm because of the aggregate impact of the content they provide access to?

TikTok recommends that the Commission should adopt a principles based approach (supplemented by guidance if necessary) thus allowing for platforms like TikTok to iterate our methodologies and technological solutions.

We welcome the opportunity to provide the Commission with information on our approach to this area which, we believe, reflects best practice in addressing harm that may be caused by aggregated content.

An inherent challenge of any recommendation system is ensuring the breadth of content surfaced to a viewer isn't too narrow or too repetitive. At TikTok we are intently focused on this challenge, and work to design a system that intersperses a variety of topics. For instance, viewers will generally not be served two videos in a row made by the same creator or that use the same sound, and we try to avoid showing people something they've seen before.

In addition, we work to carefully apply limits to some content that does not violate our policies, but may impact the viewing experience if viewed repeatedly, particularly when it comes to content with themes of sadness, extreme exercise or dieting, or that is sexually suggestive.

We understand that people express themselves in all sorts of ways on TikTok – including when they are feeling down or are going through a difficult life experience. We routinely hear from experts that closing the door on this expression can increase feelings of isolation and stigmatisation, and that enabling people to see how others cope with difficult emotions can be beneficial, especially for teens. With this in mind, our approach is to remove content that promotes or glorifies self-injury or our other policies, while allowing recovery or educational content, with limits on how often such recovery or educational content is eligible for recommendation.

Our systems do this by looking for repetition among themes like sadness or extreme diets, within a set of videos that are eligible for recommendation. If multiple videos with these themes are identified, they will be substituted with videos about other topics to reduce the frequency of these recommendations and create a more diverse discovery experience. This work is ongoing, and over the last year alone, we have implemented over 15 updates to improve these systems, along with expanding to support more languages.

Our trust and safety and product teams partner to drive this work, which is informed by academic literature and consultation with experts, such as the International Association for Suicide Prevention and the Digital Wellness Lab at Boston Children's Hospital. We will continue these efforts as we strive to recommend a diversity of content to enable an

enriching discovery experience. We are determined to provide both a welcoming space for self-expression and an enjoyable environment for our community.

Are there current practices which you consider to be best practice in this regard?

On TikTok, For You feeds help people discover a diversity of content, creators, communities, and products.

But we also understand that there are times when people's recommendations do not feel relevant anymore, or provide enough topical variety.

TikTok works continuously to improve our recommender systems to, not only improve our product, but also to develop new strategies to interrupt repetitive patterns that may include harmful content:

- Our recommendation system works to intersperse recommendations that might fall outside people's expressed preferences, offering an opportunity to discover new categories of content. For example, our systems will not recommend two videos in a row made by the same creator or with the same sound. Doing so enriches the viewing experience and can help promote exposure to a range of ideas and perspectives on our platform.
- Making content that is not appropriate for a broad audience ineligible for recommendation into For You feeds.
- Minimising recommendations of topics that could have a negative impact if viewed repeatedly. For example, topics related to dieting, extreme fitness, sadness, and other well-being topics. We also test ways to recognise if our system may inadvertently be recommending a narrower range of content to a viewer.
- Filtering out content with complex or mature themes from teen accounts, powered by our Content Levels system.
- TikTok has introduced a ['refresh' feature](#) that enables people to refresh their For You feed if their recommendations. When enabled, this feature allows someone to view content on their For You feed as if they just signed up for TikTok. Our recommendation system will then begin to surface more content based on new interactions. This feature adds to a number of content controls our community already has to shape their experience. For example, people can choose to automatically filter out videos that use specific hashtags or phrases from their For You feeds, and say "not interested" to skip future videos from a particular creator or that use a particular sound. Users can also learn [why a video is recommended](#) for them. Enabling refresh will not override any settings a user has already chosen to enable or impact accounts they have followed.

	<ul style="list-style-type: none"> As a result of our continued testing and efforts, we constantly improve our platform viewing experience so viewers now see fewer videos about these topics at a time. We are also working to recognise if our system may inadvertently be recommending only very limited types of content that, though not violative of our policies, could have a negative effect if that is the majority of what someone watches, such as content about loneliness or weight loss. Our goal is for each person's For You feed to feature a breadth of content, creators, and topics. This work is being <u>informed</u> by ongoing conversations with experts across medicine, clinical psychology, and AI ethics, members of our Content Advisory Council, and our community. <p>In addition, as the Commission will be aware, VLOPs under the DSA are required to introduce at least one option for each of their recommender systems which is not based on profiling under Article 38 of the DSA. TikTok has worked to implement this requirement and users are able to access recommended content on the service that is not based on profiling.</p>
<p>Question 21:</p> <p>Do you have any views on how requirements for commercial content arranged by a VSPS provider itself should be reflected in the Code?</p>	<p>In line with our general comments, we believe the Code in this area should be high-level and risk-based and, as to these requirements specifically, follow the position as set out in the AVMSD.</p> <p>TikTok notes here that all ads on TikTok (representing commercial communications marketed, sold or arranged by TikTok for the purposes of AVMSD) are required to comply with TikTok's Community Guidelines (as explained above) and Ad Policies.</p> <p>TikTok's Ad Policies prohibit advertisements for a wide variety of products and industries either globally (for instance, bans on gambling and tobacco products) or on a regional basis (for example, prohibiting any advertising of alcohol products to be delivered to users within the EU). In other cases, the Policies restrict the target audience for advertising certain products or services (for instance, advertising for energy drinks can only be delivered if targeted at users aged 18 and over). All ads must also comply with stringent editorial rules. These requirements reflect, and in many cases go beyond, the restrictions imposed under Article 9(1) of AVMSD and other local law obligations.</p> <p>The Community Guidelines and Ad Policies are enforced using a combination of automated and human moderation.</p>
<p>Question 22:</p>	<p>TikTok recognises that Member States are required to establish the necessary mechanisms to assess the appropriateness of the measures taken by VSPSs under Article 28b(3) of the AVMSD and that the Commission has been entrusted with this regulatory function in respect of Ireland.</p>

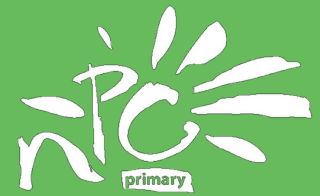
<p>What compliance monitoring and reporting arrangements should we include in the Code?</p>	<p>As regards the compliance monitoring and reporting arrangements that the Commission should include within the Code, we note that VSPSs will be/are already subject to significant transparency and reporting obligations under the DSA and under the COPD. To avoid the introduction of duplicative and burdensome reporting requirements, we would encourage the Commission to consider the extent to which existing transparency and reporting obligations under the DSA might also be able to assist the Commission in assessing and monitoring compliance with the requirements of the Code. In this way, the Commission would be able to maximise the potential for synergies in how platforms comply with it and the DSA. We respectfully suggest that - in order to make the outcomes of reporting achievable and intelligible- the Commission should set itself a high bar in any decision to deviate from these existing standards.</p> <p>However, if the Commission considers that it requires additional information in order to monitor compliance with the Code (beyond the information that VSPSs are required to make available under other regulatory regimes) we would suggest that the Commission seek to ensure any reporting arrangements under the Code are proportionate and target information which is limited to that which is otherwise necessary in this specific Code context.</p>
<p>Question 23:</p> <p>Should the Code have a transition period or transition periods for specific issues? Which areas touched on in this Call for Inputs may VSPS providers require time to transition the most? What time frame would be reasonable for a transition period?</p>	<p>We agree with the suggestion that the Code should have an appropriate transition period. Given the issues addressed by the Code are necessarily intertwined and our view that the Code should set out high level principles, we believe this transition period should apply to the Code in its entirety. At this stage, without the benefit of knowing the details of the Code, we are not in a position to provide more specific guidance or assistance to the Commission. However, once the Code is close to completion we would recommend the Commission engage with industry on this issue.</p> <p>In light of the fact that providers will have undergone and are still undergoing a very significant period of transition to ensure DSA compliance, we would suggest a minimum transition period of 12 months.</p> <p>We note that the DSA allowed a 15 month transition period for most in-scope providers and this was in circumstances where many providers had commenced DSA compliance projects long before the DSA became law. A significant transitional period will undoubtedly be required here as providers will not be aware of what obligations (and the precise nature and extent of them) will be contained in the Code until it is published.</p>

National Parents Council

Primary

Submission to

Coimisiún na Meán



Supporting Parents
Supporting Children

Developing Ireland's First Binding Online Safety Code for Video-Sharing Platform Services



Submitted
September 2023

Contents

Introduction..... 1

Submission..... 2

Appendix 1 - Parents' survey results 14

Appendix 2 - Children's survey results 24

Appendix 3 - Parent’s comments - responses to question 21 32

Introduction

The National Parents Council Primary welcomes the opportunity to submit its views to Coimisiún na Meán regarding the Online Safety Code for Video-Sharing Platform Services

NPC is the only recognised representative organisation for parents in education in Ireland. NPC was established as a charitable organisation in 1985, under the programme for Government, as the representative organisation for parents of children attending primary school. It received statutory recognition in the Education Act 1998.

NPC Vision

NPC want to see an Ireland where **every** child has the opportunity to reach their full potential.

NPC Mission

NPC exists to ensure that all parents are supported and empowered to become effective partners in their children's education. NPC will work to increase the capacity and capability of the education sector, to achieve true partnership and deliver better outcomes for all children.

NPC's Key Activities are:

- Representing the parents' voice in primary education
- Advocacy
- Building participation
- Service delivery

NPC Service Delivery

NPC services are aimed at empowering parents so that they can support their children in all aspects of education.

Helpline

The NPC helpline is a national confidential service for parents. The helpline staff listen and give information and support to parents to help them make the best possible decisions for and with their children.

Training and Development

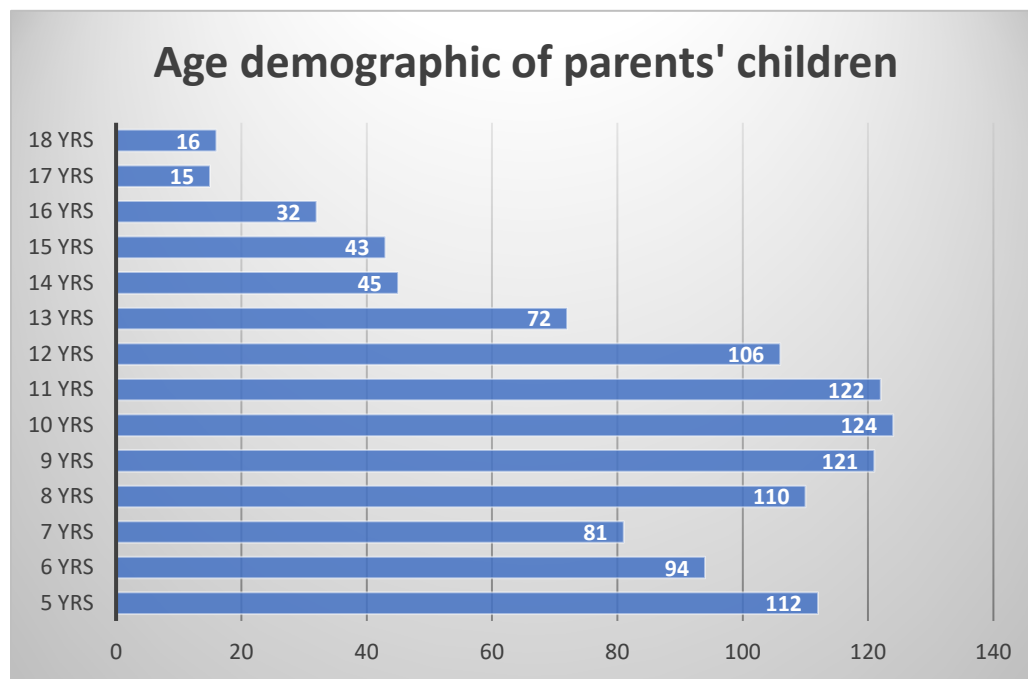
The NPC Training and Development programme is a national programme of training, development and support for parents. The purpose is to empower parents to play an active part in their child's education at every level.

Website

The NPC's website www.npc.ie aims to provide parents with information regarding education. The site also allows parents an opportunity to give NPC their views regarding education issues.

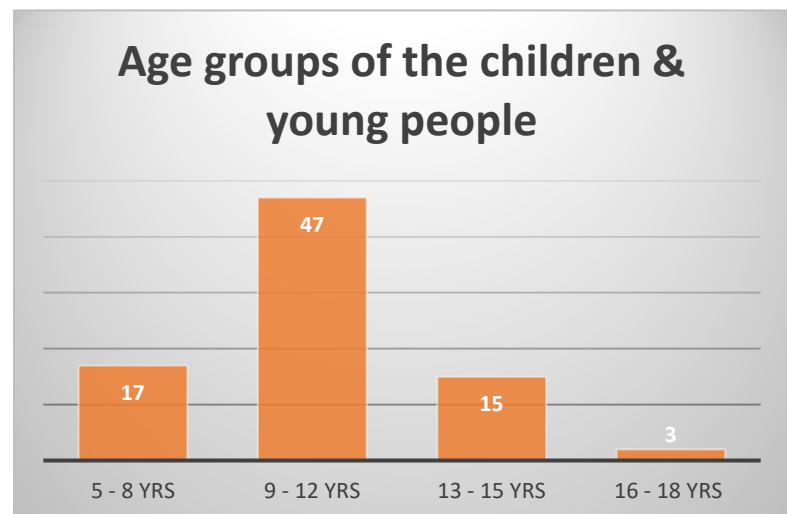
NPC Submission

To inform the National Parents Council Primary's submission on developing Ireland's first binding online safety code for video-sharing platform services, two online surveys were developed in order to hear parents' opinions and the opinions of their children. The surveys asked for feedback on their and their child's experience of consuming video content as well as their views on important issues in the development of the code. The two surveys were sent to NPC members and those on the NPC contacts database who have consented to taking part in surveys. Links to the surveys were also displayed on our website (www.npc.ie) and on the NPC social media platforms. The survey ran from 1pm on 28th August 2023 to 12 midnight on 30th August 2023.



There were a total of 595 responses from parents who were asked to list the number of children and their ages, representing a total of 1,093 children aged between 5 years and 18 years.

82 children and young people aged between 5 and 18 years of age responded to the survey, they were asked to indicate to which age group they belonged, this chart represents their responses.



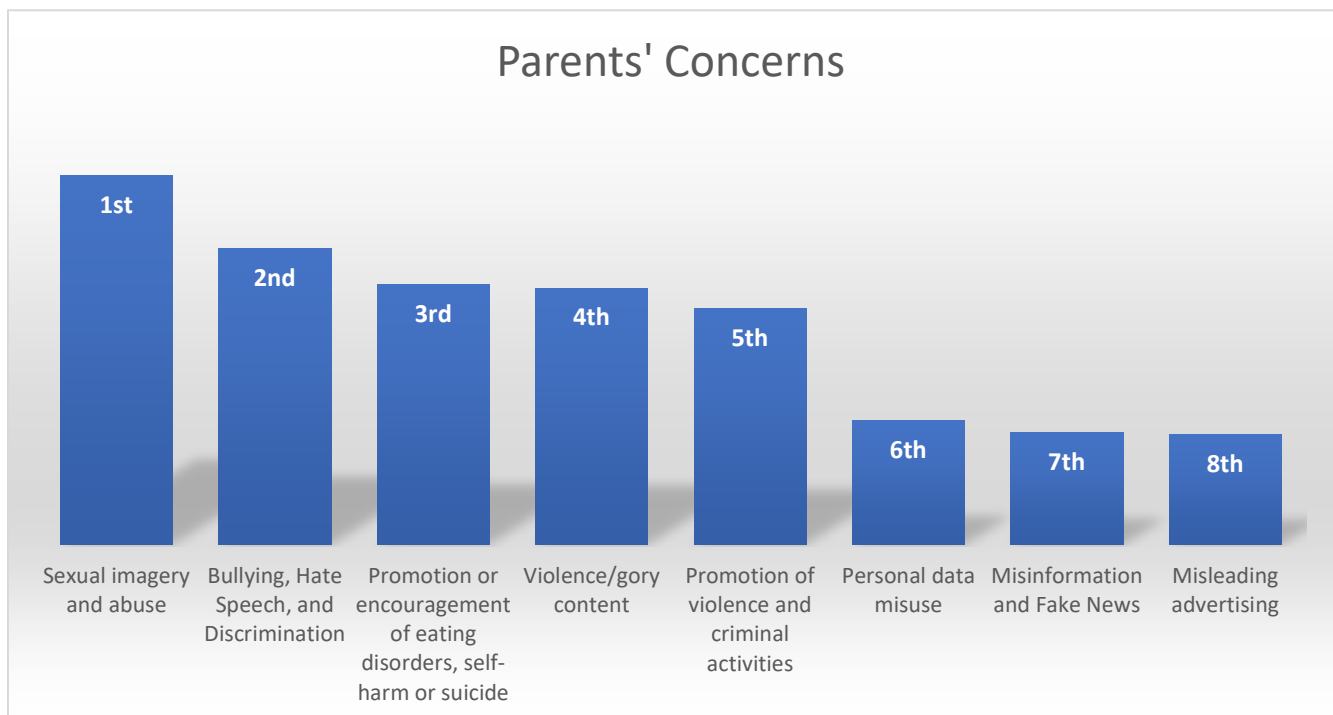
We asked parents for their views on their understanding of online safety and the regulation of video-sharing platform services (VSPS) in Ireland. With the growing importance of the internet and video-sharing websites in the lives of our young people, and the many benefits it provides, it is vital to ensure the safety of users, especially children, while they engage with online content.

The 'Call for Inputs' document gave a set of guideline questions, the survey results do not address all of the guideline questions, the survey only asked for opinions on the questions that we felt were relevant to parents, children and young people.

Question 1: What do you think our main priorities and objectives should be in the first binding Online Safety code for VSPS? What are the main online harms you would like to see it address and why?

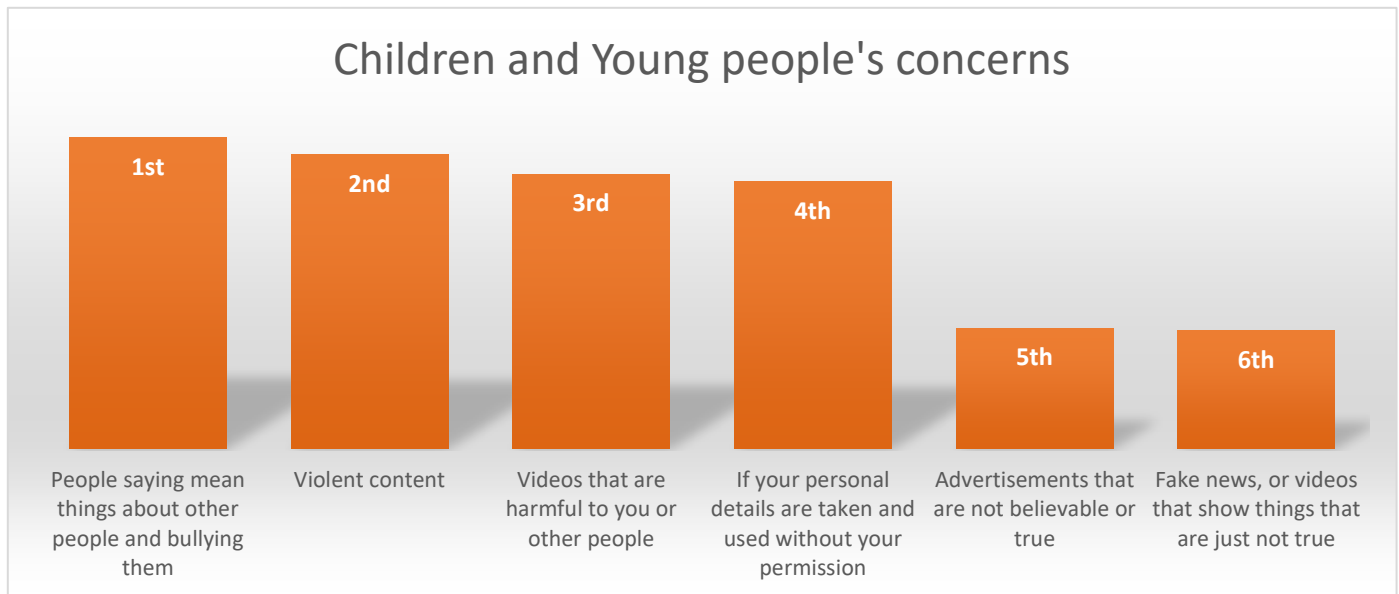
Question 2: What types of online harms do you think should attract the most stringent risk mitigation measures by VSPS? How could we evaluate the impact of different types of harms e.g. severity, speed at which harm may be caused? Is there a way of classifying harmful content that you consider it would be useful for us to use?

In order to ensure that as many parents, children and young people could participate and have their voices heard on this topic, the questions in this survey were asked in accessible language, and although the questions asked may not sit neatly under the guide questions asked in the 'Call for Inputs' document issued from Coimisiún na Meán, it is hoped that the responses from the survey will give insight into the areas asked for in the document. With that in mind, the following addresses both questions 1 and 2.



Parents were asked what types of online risks concerned them the most, they were asked to rank the options given from the most concerning to the least concerning.

Weighted values were applied to give ranked results and the number one concern of parents was sexual imagery and abuse, **72%** of parents gave this as the most concerning online risk.

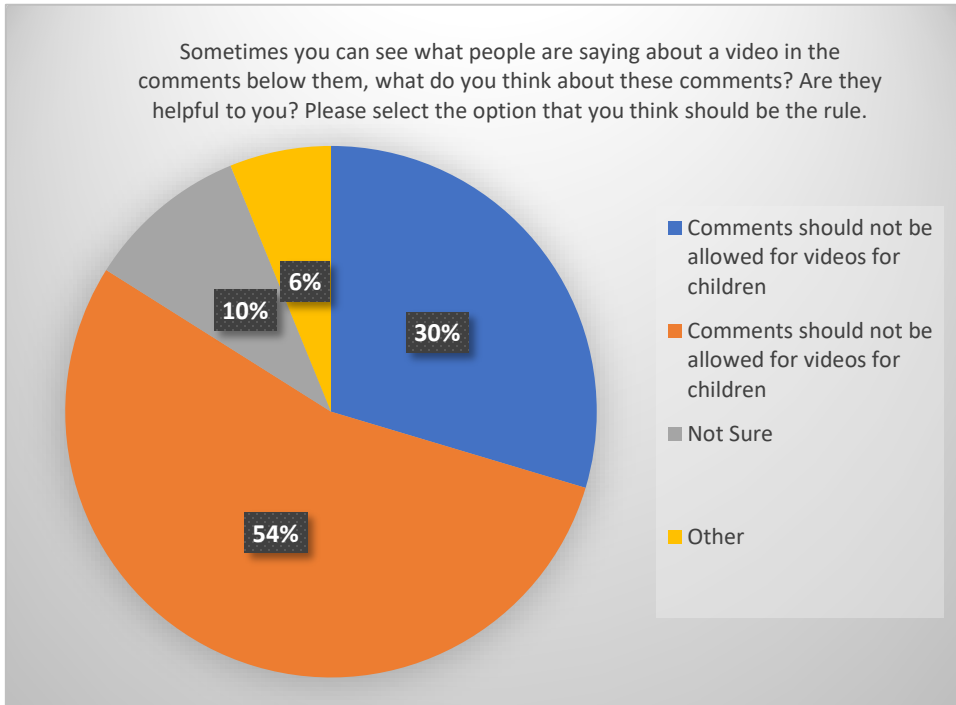


Although the weighted responses from children and young people put 'People saying mean things about other people and bullying them' as their number one concern, interestingly **30%** of them gave their number one vote to 'If your personal details are taken and used without your permission' as opposed to **28%** giving their number one vote to 'People saying mean things about other people and bullying them'.

Question 7: To what extent, if at all, should the Code require VSPS providers to take measures to address content connected to video content?

Comment sections connected to videos shared online are often filled with negative and toxic comments, including insults, threats, and arguments.

Parents were asked who they thought should be responsible for regulating the content connected to videos shared online, in particular the comments associated with the videos. **70%** of parents thought that comments should be disabled for videos aimed at children, and **22%** felt that the comments should be effectively monitored. The remainder of parents were unsure how they felt about this.



54% of the young people surveyed felt that comments should be allowed but they should be monitored.

Comment from a young person:

“Comments could be fake and written by person(s) who posted the video. Also comments can be abusive and kids read these comments and think its ok to say that stuff.”

Question 8: How should we ask VSPS providers to introduce a feature that allows users to declare when videos contain advertising or other types of commercial communications? Should the Code include specific requirements about the form in which the declaration should take? What current examples are there that you regard as best practice?

Clearly labelling sponsored content in videos aimed at children is essential for transparency, it helps children and their parents understand that what they are watching is a form of advertising rather than regular content. Declaring sponsored content allows viewers, including children, to make informed decisions about the content they engage with. It helps them distinguish between organic content and promotional material. By clearly marking sponsored content, video platforms could also use this as an educational opportunity to teach children about advertising and the difference between regular content and advertisements.

Parents were asked if they thought sponsored content should be clearly labelled and regulated to ensure that children can distinguish between regular content and advertisements, or if they believed that sponsored content should not feature at all in videos aimed at children and such content should be completely separate from videos meant for young audiences.

85% of parents believed that sponsored content had no place in videos aimed at children.

One parent commented:

"There should be no advertising whatsoever to minors online, not only things deemed generally inappropriate but also harmful to the individual or unhealthy, which varies widely from person to person. There is no way to fully monitor the damage so it should not be considered at all, it should all be banned for children."

39% of the children and young people surveyed thought that it should be very clear and obvious to them when products or services were being promoted, but **50%** felt that these promotions had no place in video content aimed at children or younger people.

One young person commented:

"They should say if their video is just really an ad to get me to buy something"

Question 9: How should we ask VSPS providers to introduce and design a flagging mechanism in the Code? How can we ensure that VSPS providers introduce the mechanism in a user-friendly and transparent way? How should we ask VSP Providers to report the decisions they've made on content after it has been flagged? To what extent should we align the Code with similar provisions on flagging in the DSA?

This question was posed to parents, children and young people in an accessible way, parents were asked firstly, if they knew they could report harmful content, and if they did, had they ever done so? They were then asked if they had been told of the outcome and if they were happy with the outcome. Similarly, children and young people were asked the same, did they know they could report something, had they done so, did they know what happened and if they did were they happy with it?

Whilst 79% of parents said they were aware of being able to report content of concern to video sharing platforms, some 48% of parents had actually done so. Of this 48%, just 22% of parents were told of the outcome, and only 9% of those were happy with the outcome. Some parents stated that they weren't sure of the outcome as they had blocked the content or simply didn't want to go back and check if it had been removed as it was just too distressing to view again.

Providing feedback to complainants on the actions taken in response to their complaints or concerns promotes transparency. It helps users understand that their concerns have been heard and addressed, which can build trust in the platform's moderation process. It can also help users to understand what types of content are considered inappropriate and what actions the platform takes to address such content, this can contribute to users' awareness and responsible online behaviour. NPC believes if sanctions for posting inappropriate content are clear, that knowing that there are consequences for posting inappropriate content may deter some users from engaging in such behavior in the first place and publicising sanctions can serve as a deterrent to potential rule violators.

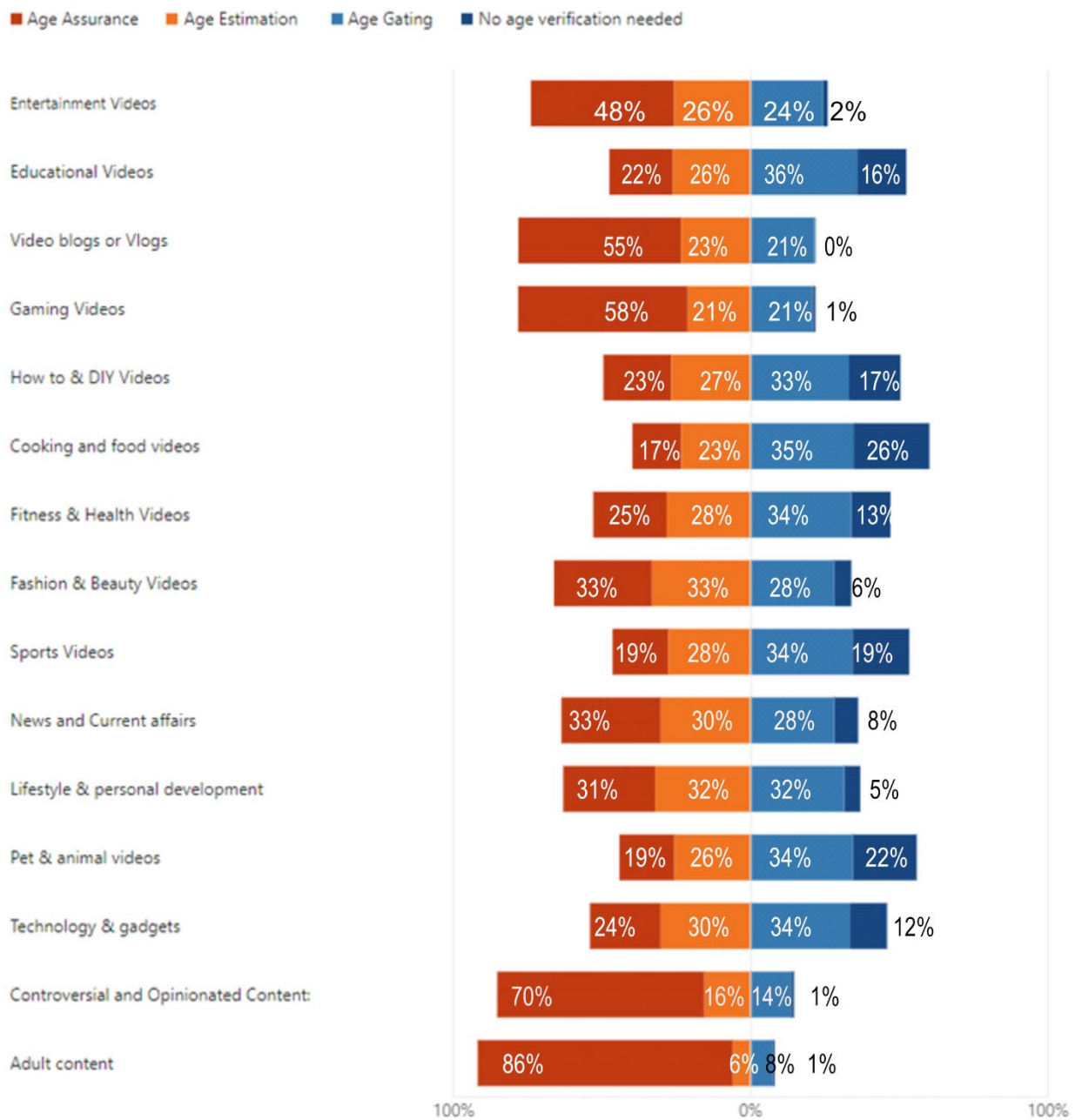
65% of young people were aware that they could report unsuitable content, and of that cohort **43%** had actually done so, but only **5%** of that **43%** had been told of the outcome.

The best interests of the child should be a primary consideration, balancing protection with the rights of the child for freedom of expression, participation, and access to information. Flagging mechanisms should be prominent, age-appropriate, straightforward and understandable for parents, children and young people to use.

Question 10: What requirements should the Code include about age verification and age assurance? What sort of content should be shown by default to users who are logged out or in private browsing mode and whose age cannot be verified or assured? What evidence is there about the effectiveness of age estimation techniques? What current practices do you regard as best practice? Where accounts are not age verified should default privacy settings be used, should content default to universal content and should contact by others be more limited?

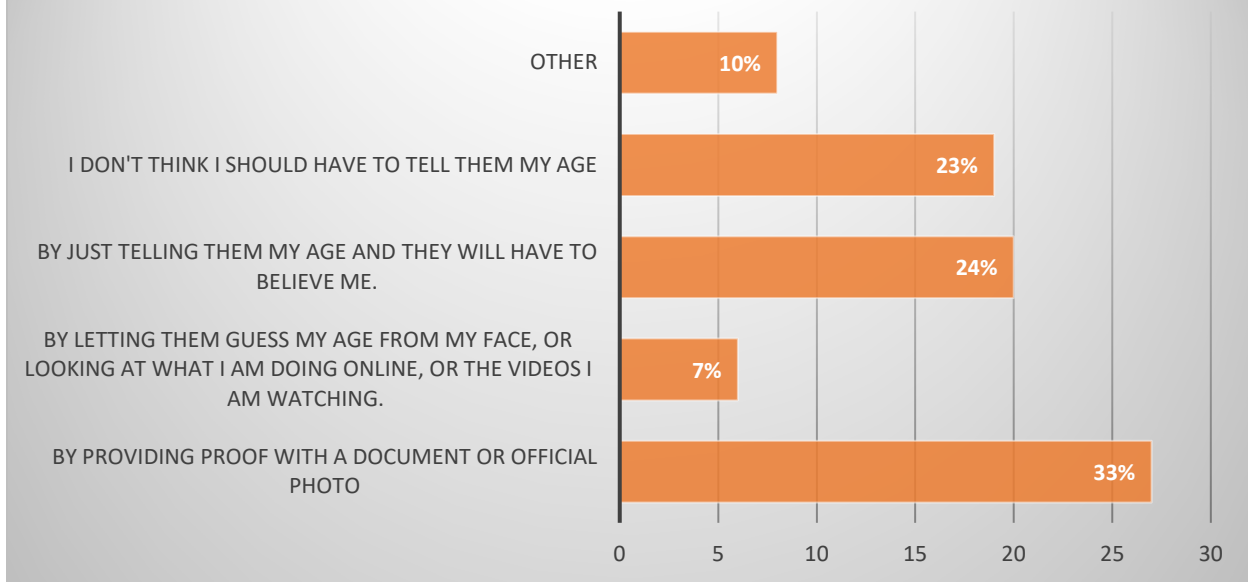
NPC's survey asked parents what types of age ratings (if any) should be applied for different video content, and the majority believed that there should be an age rating applied to most video content. Parents stated that adult, controversial and opinionated content should have an Age Assurance method to ascertain the age of the viewer, and a third of parents believed that fashion, beauty, personal development and lifestyle should have an Age Estimation method. Over a third of parents said that educational content such as DIY, cooking, fitness, sport, pets, and technology should only require Age Gating requirement.

A more detailed explanation of the results are below:



Responses from the children and young people were quite divided, **33%** of the young people felt it should be an official document, **24%** said it should be an Age Gating method and another **24%** said they should not be required to give their age.

Age Verification preferences for children & young people

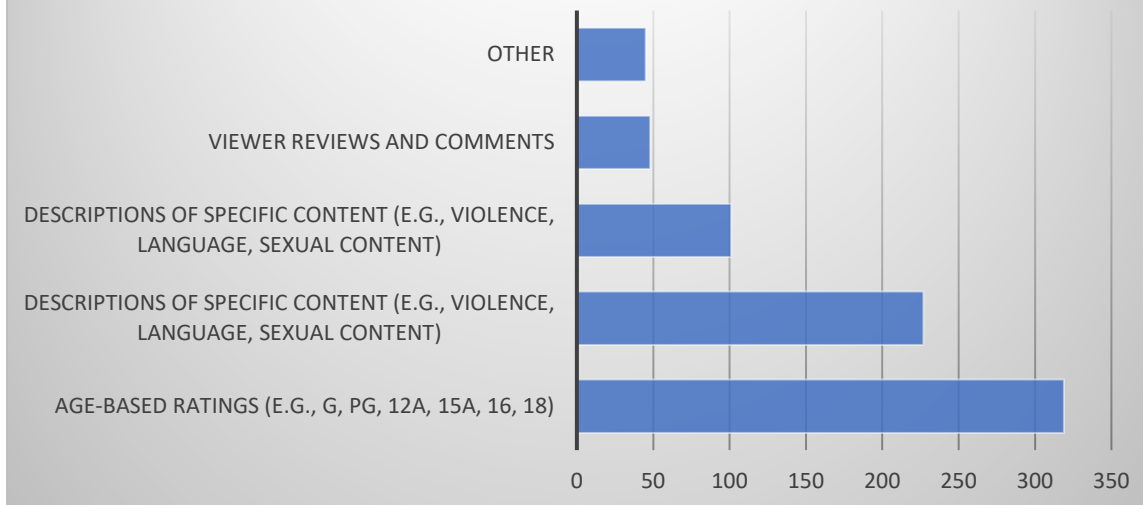


Question 11: What requirements should the Code have in relation to content rating? What do you consider to be current best practice? What experiences have you had using content rating systems on platforms and do you think they have been effective? What steps could we ask VSPS to take to ensure content is rated accurately by users?

Parents were asked if they were familiar with different systems used to rate video content, and **55%** of parents stated that they were somewhat familiar with them.

The survey then asked what type of system they thought might be most useful to them.

Parents' preferences for systems of content rating



When asked about a favoured system of rating the content, **54%** favoured a system of age rating similar to that used for cinema content as a way of ascertaining whether content was suitable for their child or not. Some parents stated that they relied on websites such as ([Common Sense Media: Age-Based Media Reviews for Families | Common Sense Media](#)) for information about content.

48% of parents were not aware of any content rating information for selecting content on video sharing platforms, and **30%** said they had only used them occasionally.

67% of parents felt that video sharing platforms did not provide enough information about their content to allow users to make informed decisions before watching them.

40% of young people said they found descriptions of the content the most useful when deciding whether to view it or not, and **39%** said the age ratings were more effective, however, the majority of them (**69%**) said they were unaware or unsure if they had seen any of the platforms with these descriptions on them.

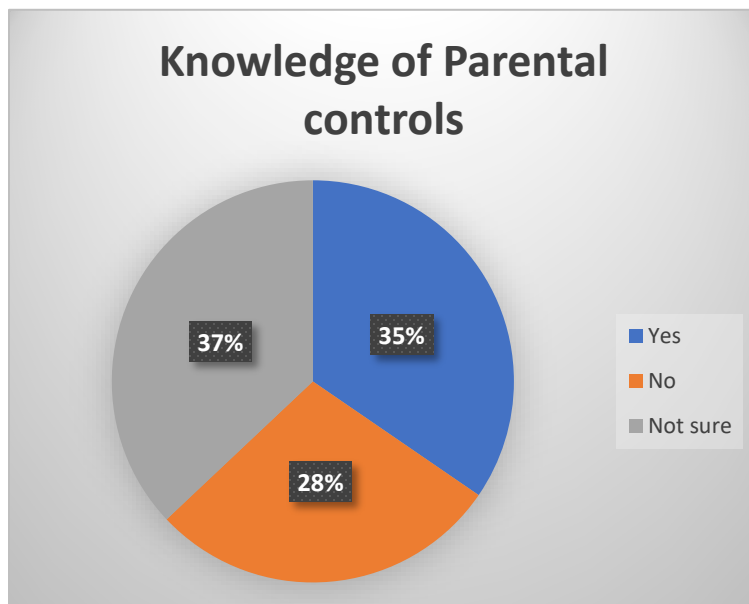
The survey asked the children and young people whether, had they have seen the descriptions, it would have changed their minds about viewing it, **57%** said that it may have. **47%** said there was not enough information provided by the platforms before they viewed the content.

Question 12: What requirements should the Code have in relation to parental control features? How can we ensure that VSPS providers introduce the mechanism in a user-friendly and transparent way? Can you point to any existing example of best practice in this area? Should parental controls be ‘turned-on’ by default for accounts of minors or where age is not verified?

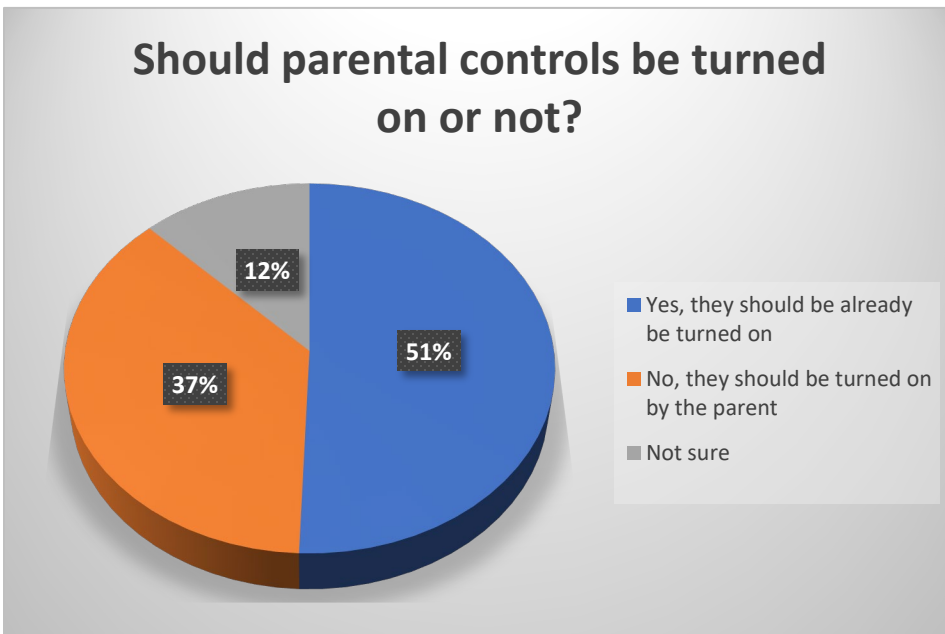
Children and teenagers often explore online content and may accidentally come across inappropriate material. Default parental controls can help mitigate accidental exposure to such content. Parental controls can provide an additional layer of protection for children and teenagers online. They can help prevent access to age-inappropriate content, limit screen time, and protect against potentially harmful interactions, they can also direct young people to more appropriate educational content that *is* suitable for them.

When asked if they were aware of parental controls, the vast majority of parents (**94%**) said they were aware or at least somewhat aware of parental controls that are available on digital devices and online platforms, with **49%** of parents saying they used them regularly, **33%** used them occasionally and **17%** said they did not use them at all.

Only **13%** of parents were confident in their ability to use parental control features to manage the content their children could access and **10%** of parents did not feel confident at all. **94%** of parents thought that parental controls should be turned on by default.



Only **35%** of young people were aware of parental controls,



When asked whether they thought that parental controls should be on all videos that are made for children and young people or should that be up to the parent to put them on, **51%** of the young people felt they should be turned on by default.

Question 17: What approach do you think the Code should take to ensure that the safety measures we ask VSPS providers to take are accessible to people with disabilities?

If accessibility is to be integrated as a safe user experience, it needs to encourage the adoption of inclusive design principles from the early stages of platform development to ensure accessibility is integrated into the user experience. Video sharing platform providers can create a more inclusive and accessible online environment for individuals with disabilities by implementing safety measures, ensuring that they can fully participate in the digital world, such as:

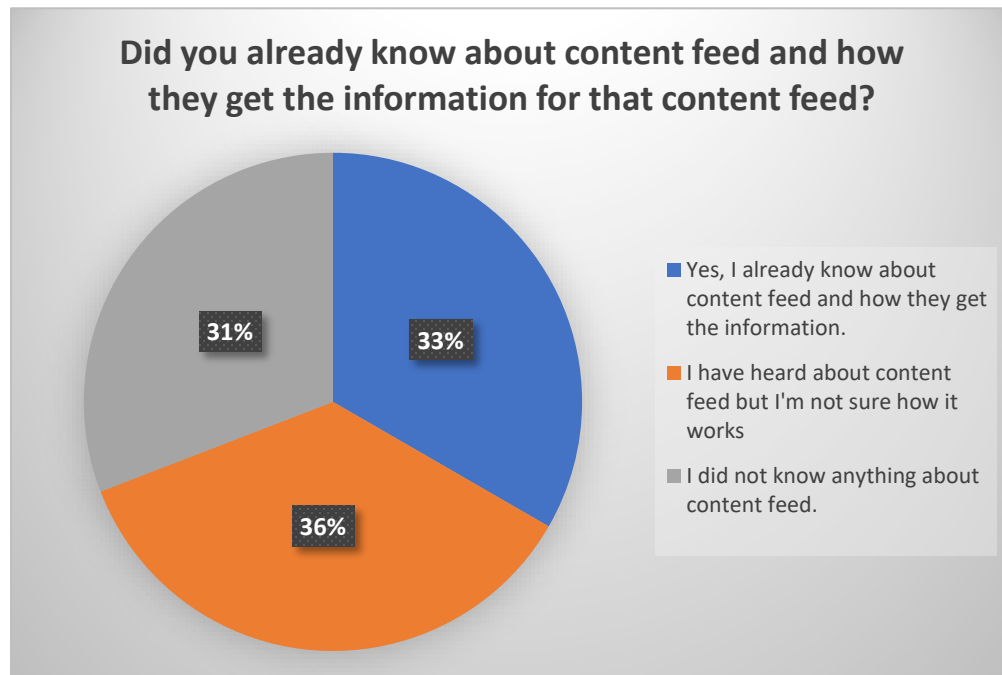
- Closed captioning
- Audio description
- Accessible play controls
- Transcripts of video
- Alternative formats
- Content guidelines

It is vitally important that providers collaborate with advocacy groups that represent people with disabilities to gain insights, feedback and expertise in improving accessibility safety features , as well as evolving best practices.

Question 20: What approach do you think we should take in the Code to address feeds which cause harm because of the aggregate impact of the content they provide access to? Are there current practices which you consider to be best practice in this regard?

Content feed algorithms, while maybe designed to provide personalised and engaging online experiences, can pose several potential risks for children and young people. Algorithms may inadvertently expose children and young people to age-inappropriate content, including violence, explicit material, or harmful ideologies. These algorithms often base recommendations on user behaviour, which can lead to unexpected and unsuitable content appearing in feeds. Platform providers should prioritise the well-being and safety of young users when designing and implementing content algorithms.

Whilst **72%** of parents said that they were well aware of content feed and advertisements associated with video content were different from person to person based on their online activity, children and young people were not quite as aware.



When asked about their knowledge of content feed and how it works, only **33%** of young people were aware of it and how it works.

Conclusion

The National Parents Council Primary welcomes the focus of the Coimisiún na Meán on the development of an Online Safety Code for Video-Sharing Platform Services. We believe that the level of engagement from parents and their children over a short consultation period by NPC, shows the importance of this issue to them. We look forward to our further engagements with Coimisiún na Meán on this and other matters and we will continue to promote the development of opportunities to hear the opinions of parents and those of their children on these important issues leading to a safer and more enriching online experience.

Parents' survey results

Keeping our children safe and informed when watching online videos

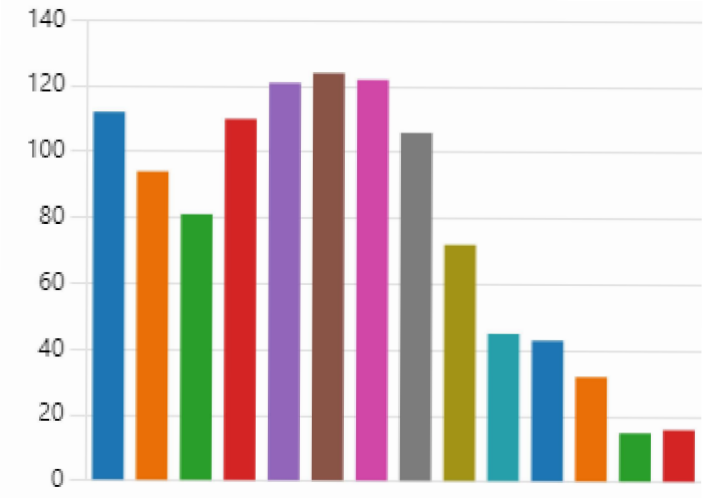
595
Responses

22:37
Average time to complete

Closed
Status

1. It would really help us if you could tell us how old your children are, please tick all that apply.

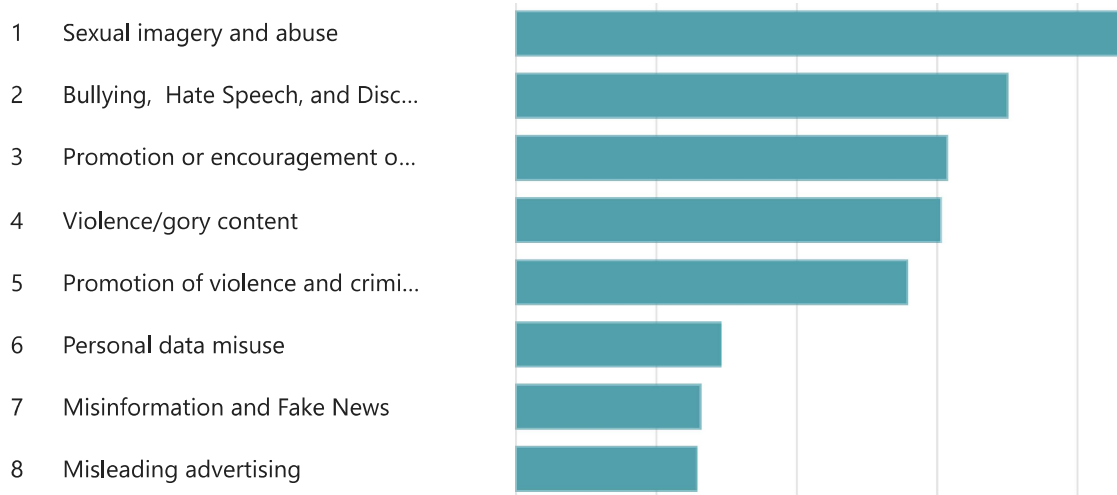
5yrs	112
6yrs	94
7yrs	81
8yrs	110
9yrs	121
10yrs	124
11yrs	122
12yrs	106
13yrs	72
14yrs	45
15yrs	43
16yrs	32
17yrs	15
18yrs	16



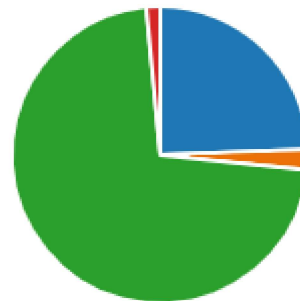
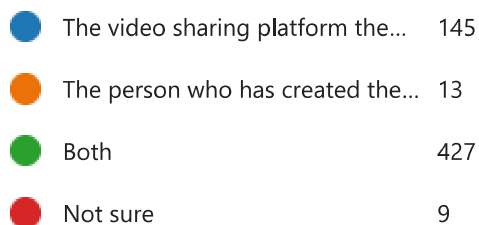
2. When our children are viewing video content online, we would hope that they are finding the content enjoyable and educational, but sometimes their experiences may have a negative impact on them.

In your opinion, what types of online content would concern you the most?

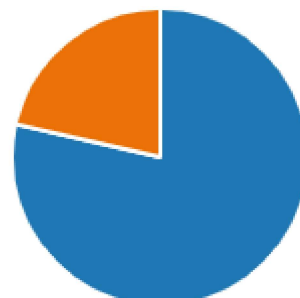
Please drag the options to indicate (in your opinion) the most concerning to the least concerning



3. Who do you think should be responsible for regulating content on videos shared online?

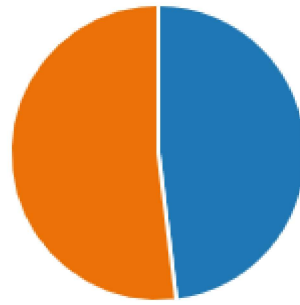


4. Did you know you can report harmful content on video sharing platforms?



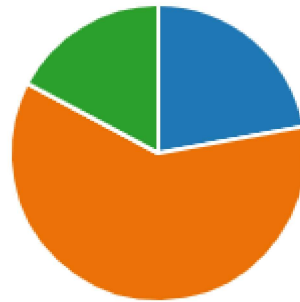
5. If yes, have you ever reported anything?

● Yes	223
● No	241



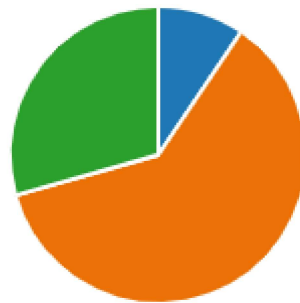
6. Were you told of the outcome?

● Yes	72
● No	197
● Other	56

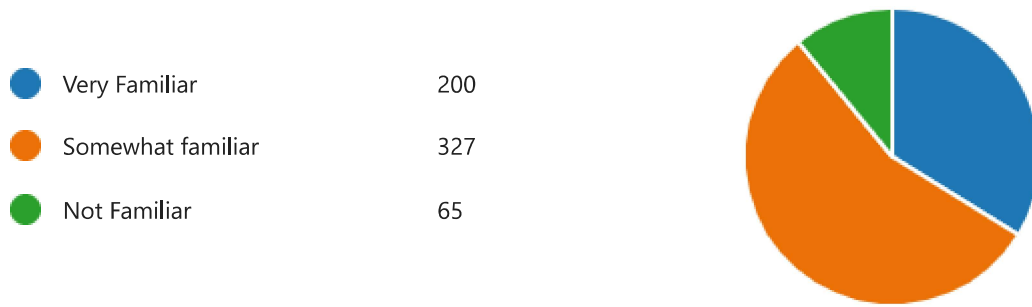


7. Were you happy with the outcome?

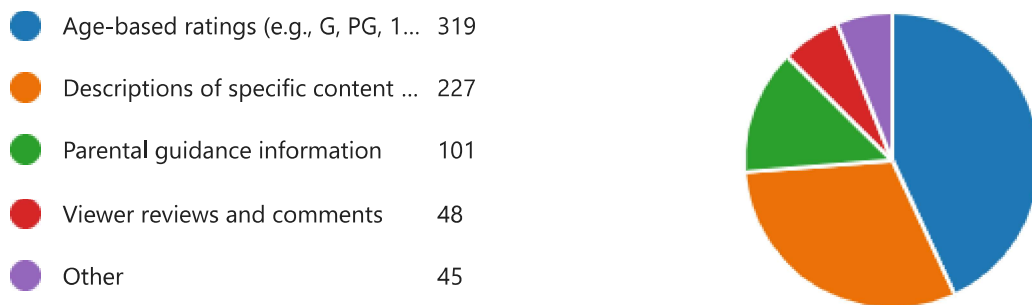
● Yes	28
● No	181
● Other	87



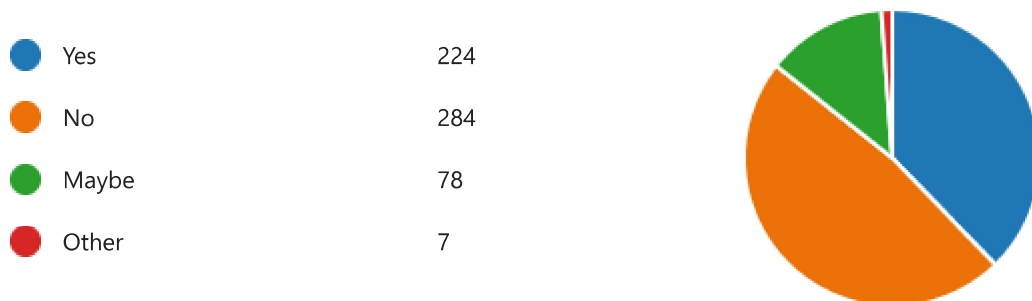
8. How familiar are you with content rating systems used in movies, television, streaming services, and video games?



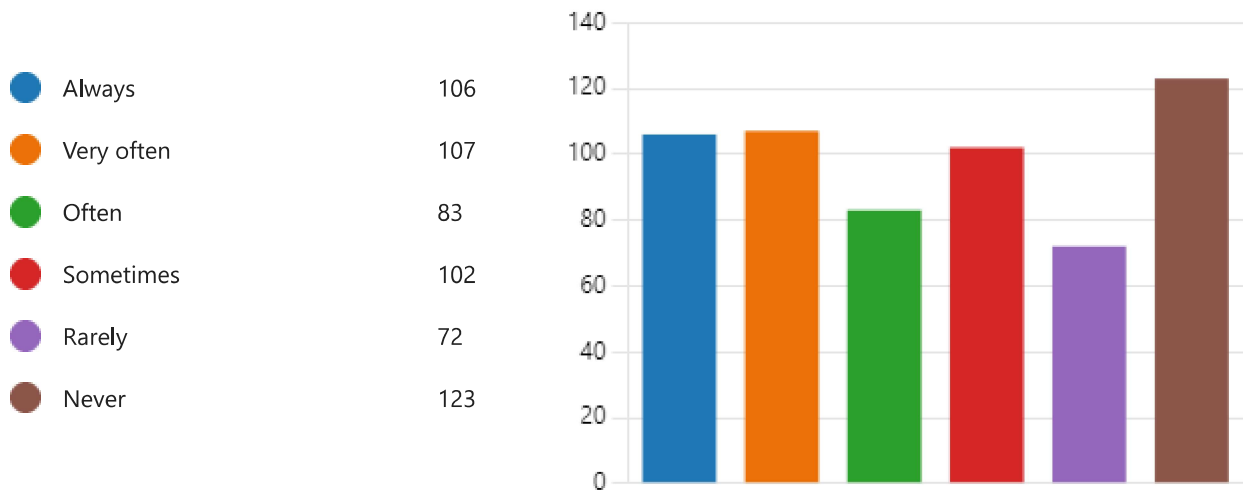
9. What types of content rating do you think are most helpful in deciding if a content is okay to watch for your children? (Please select all that apply)



10. Are you aware that you can use content rating information for selecting content on different platforms (Such as YouTube, TikTok, Video Games, Instagram)?

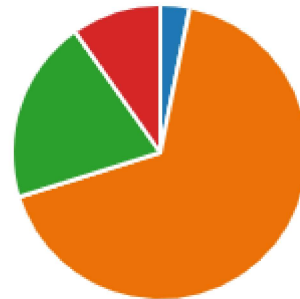


11. How frequently do you use content rating information when selecting videos or monitoring your children's viewing on Video Sharing Platform Services?



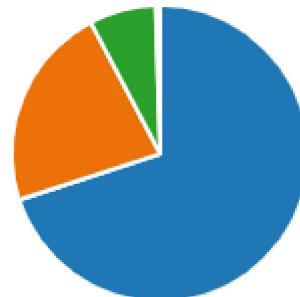
12. Do you believe that video sharing platforms provide enough information about their content to allow users to make informed decisions before watching them?

Yes	19
No	397
Not sure	119
I am unaware of any content rat...	58



13. How do you feel about comments linked to videos intended for children? Please select the option that best represents your viewpoint:

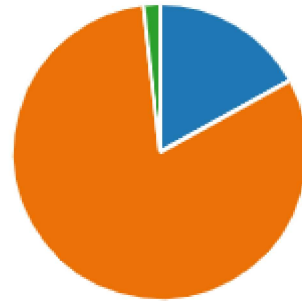
Comments should be disabled f...	415
Comments should be effectively...	132
Unsure	43
Other	3



14. How do you feel about sponsored content (advertisements or promotions) that is integrated into videos aimed at children on platforms like TikTok, Instagram, and YouTube?

Please select the option that best represents your viewpoint:

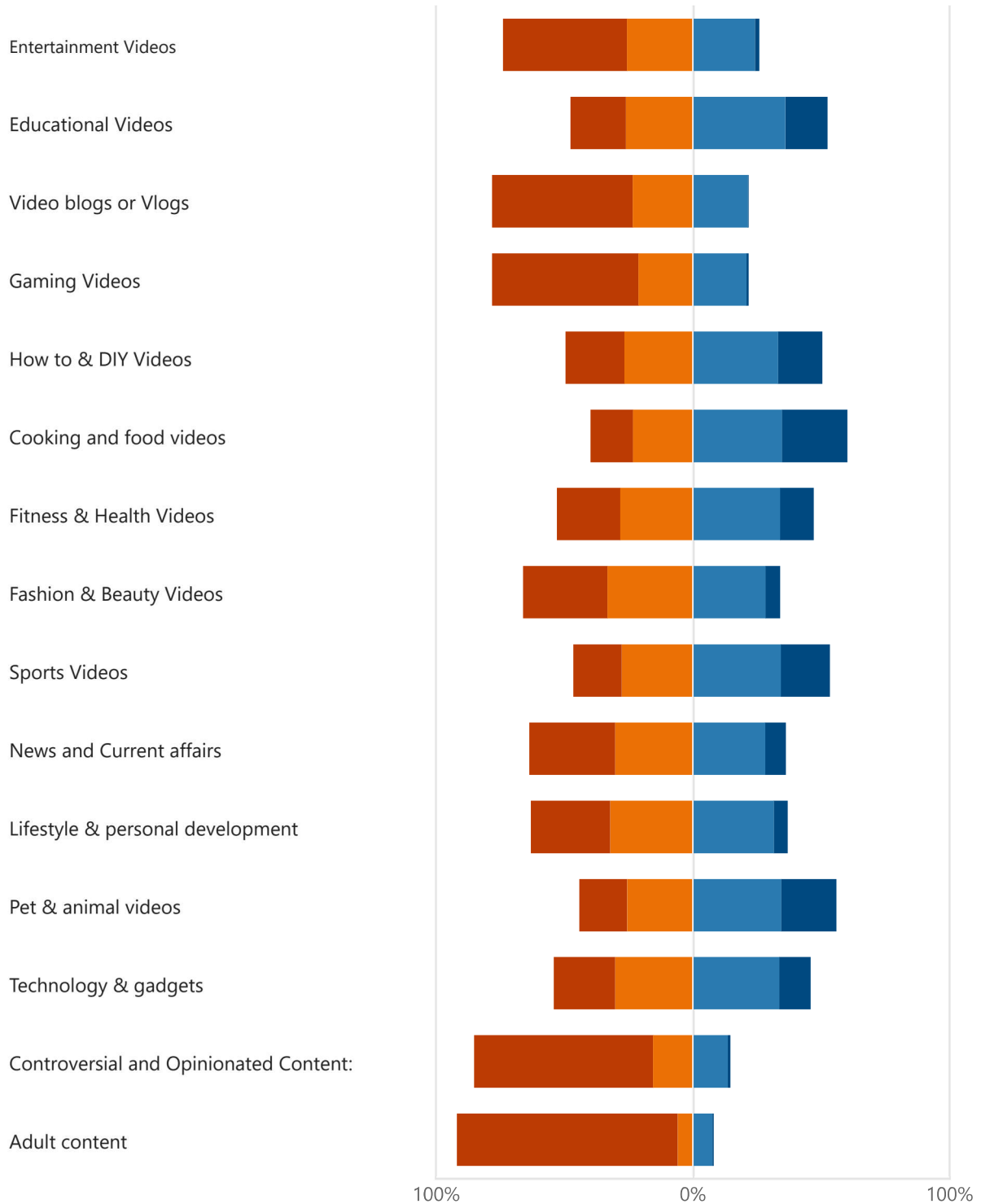
- I believe that sponsored content... 101
- I believe that sponsored content... 482
- Unsure 11
- Other 0



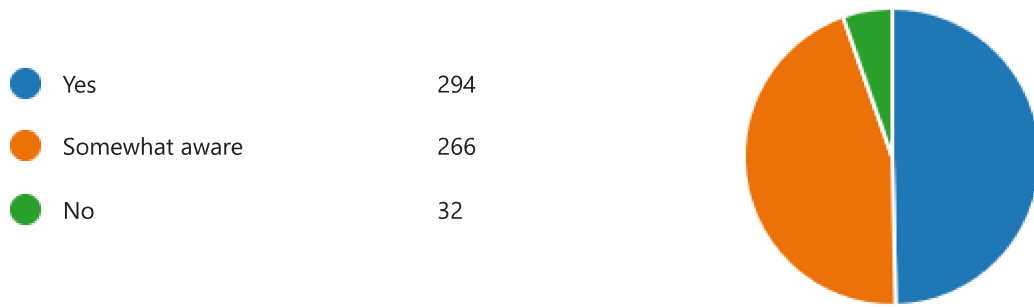
15. Below are some examples of online activities.

Which age verification method would you prefer to safeguard your child for each online activity?

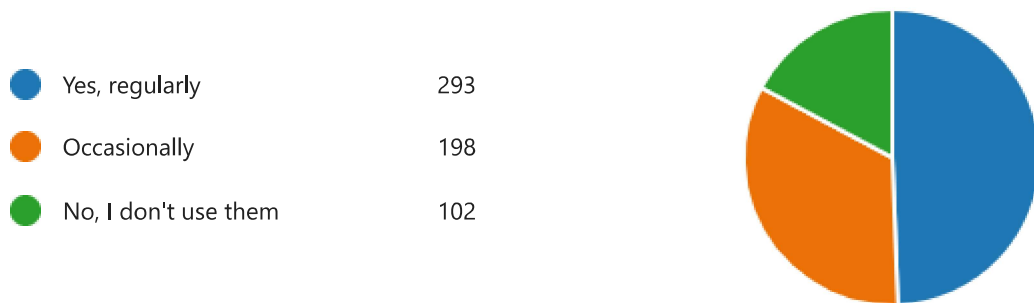
■ Age Assurance
 ■ Age Estimation
 ■ Age Gating
 ■ No age verification needed



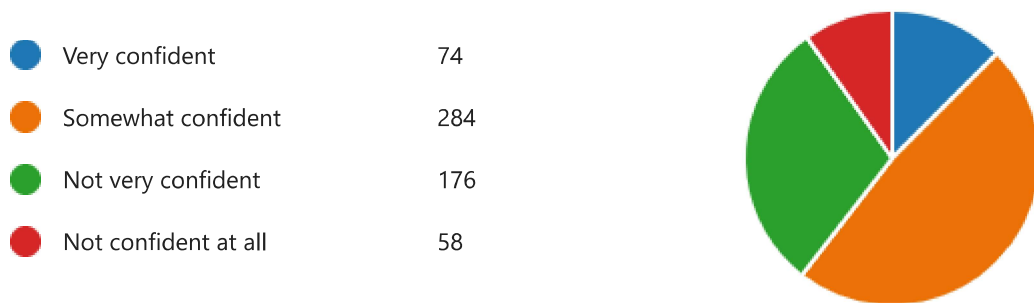
16. Are you aware of parental control features available on digital devices and online platforms?



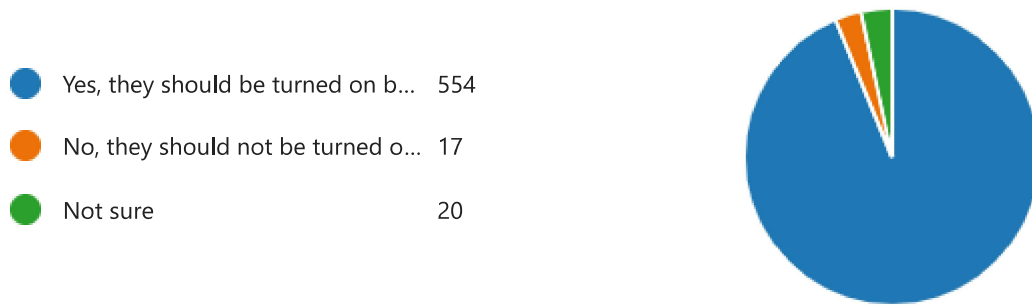
17. Do you currently use parental control tools to monitor and regulate your child's digital usage?



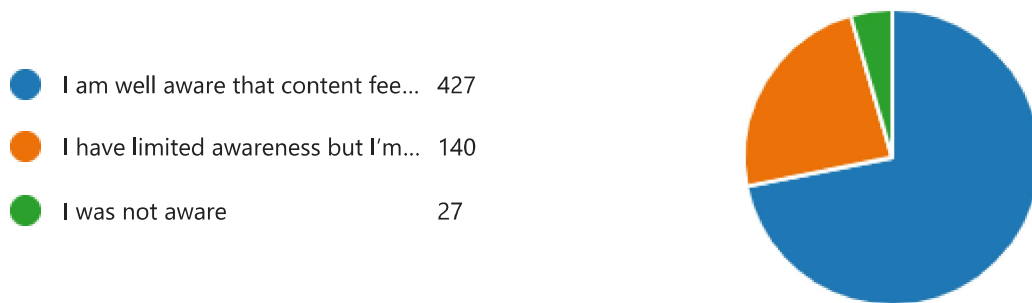
18. How confident are you in your ability to use parental control features to manage what content your children can access?



19. In your opinion, should parental controls be 'turned-on' by default for accounts of minors or where age is not verified?



20. Are you aware that the **content feed** and **advertisements** associated with video content are different from person to person based on their online activity?



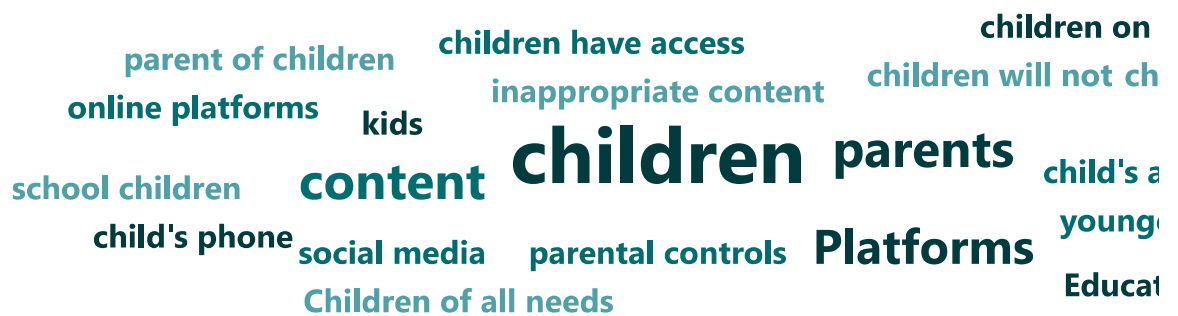
21. Keeping in mind the above survey, is there anything else the Commission for Online Safety could do to support children and young people with additional needs using video sharing platforms (Such as Tiktok, Instagram, YouTube)? (Parental Controls, Accessibility, Online Harms etc)

277

Responses

Latest Responses

113 respondents (41%) answered **children** for this question.



Children's survey results

Your safety and enjoyment when watching online videos

82

Responses

23:43

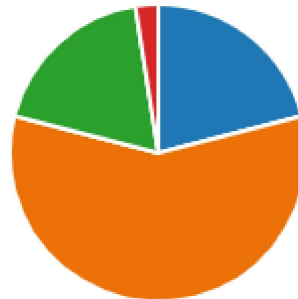
Average time to complete

Closed

Status

1. Knowing how old you are will help us with the survey, can you tell us what age range you fit into?

● 5yrs - 8yrs	17
● 9yrs - 12yrs	47
● 13yrs - 15yrs	15
● 16yrs - 18yrs	2

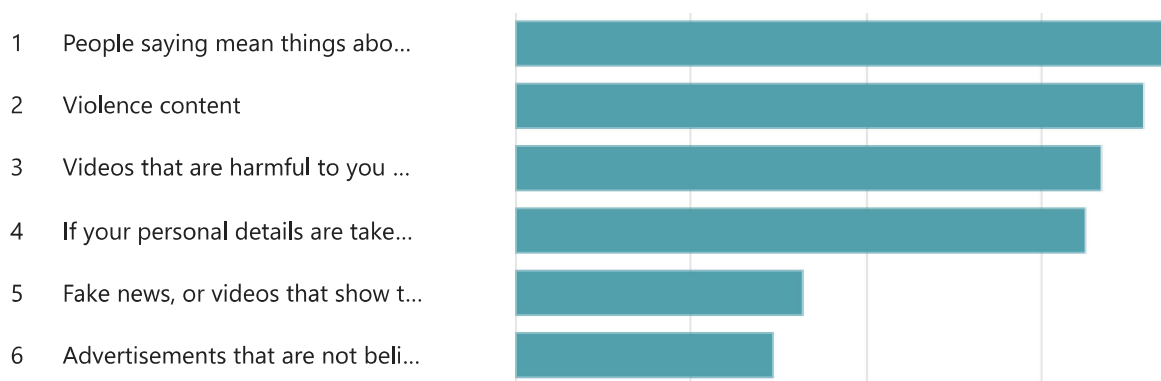


2. When we watch videos online, it is usually for fun, enjoyment and or to find out about something or learn how to do something. But sometimes we can come across videos that aren't nice, they can be frightening, confusing or they can make us feel bad about ourselves.

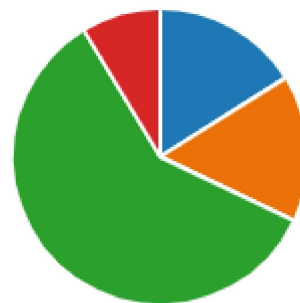
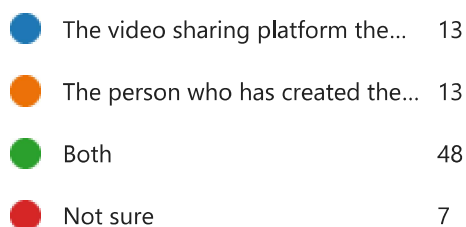
Sometimes it can be hard to know if the person is telling the truth or are they just getting paid to say something?

We would like to know what you think are the worst kind of videos or videos that you think children or young people should not be looking at.

Below you will see different kinds of video content, can you let us know what you think are the worst types by putting them in order? You can drag and drop the different types - putting the worst ones first and the ones that don't bother you as much at the bottom.

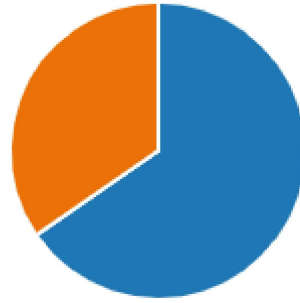


3. Who do you think should be in control or responsible for the types of videos shared online?



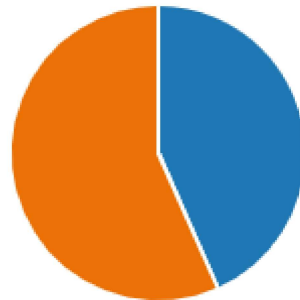
4. Did you know you can report a video if you think it might be bad or unsuitable in some way?

● Yes	53
● No	28



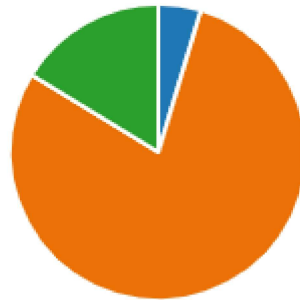
5. If yes, have you ever reported anything?

● Yes	23
● No	30



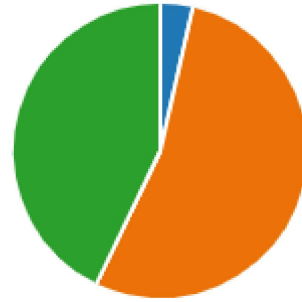
6. Did you find out what happened?

● Yes	2
● No	34
● Other	7



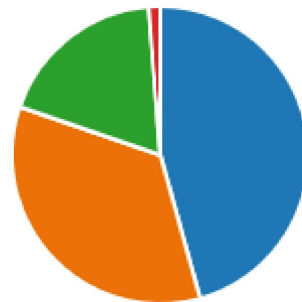
7. If you did find out what happened -were you happy with what happened?

● Yes	1
● No	15
● Other	12



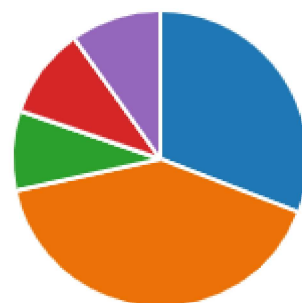
8. Have you ever seen videos that describe what is in the video or who the video is for before you watch it?

● Yes	37
● No	28
● Maybe	15
● Other	1



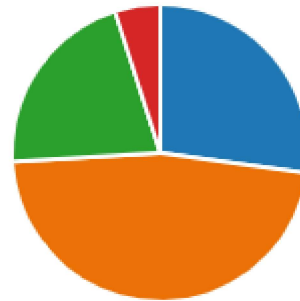
9. If you wanted to know what to expect in a video before you watched it or wanted to know if it was suitable for you to see, what kind of things would help you decide?

● Age-ratings like they use in the ...	25
● A description saying things like: ...	33
● Emoji's, star ratings, thumbs up ...	7
● People's comments on the video	8
● Other	8



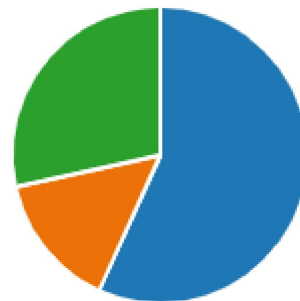
10. Did you know that video sharing platforms like YouTube, TikTok, video games and Instagram already have these descriptions on them?

● Yes	22
● No	38
● Not sure	17
● Other	4



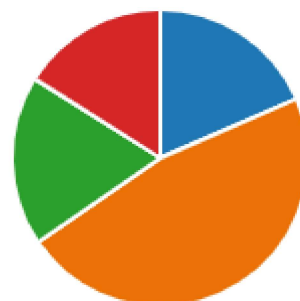
11. If you see or have already seen these kinds of descriptions telling you what might be in the videos before you look at them, do you think it would make you change your mind about watching something?

● Yes	46
● No	12
● Maybe	23



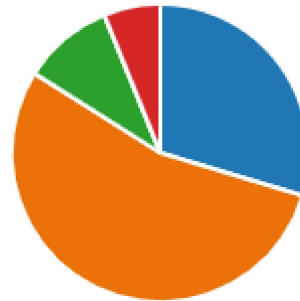
12. Do you think that there is enough information or descriptions about what is in videos before you watch them?

● Yes	15
● No	38
● Not sure	15
● I have never seen any of these d...	13



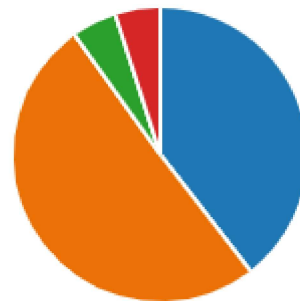
13. Sometimes you can see what people are saying about a video in the comments below them, what do you think about these comments? Are they helpful to you? Please select the option that you think should be the rule.

● Comments should not be allow...	24
● Comments should be allowed b...	44
● I'm not sure	8
● Other	5



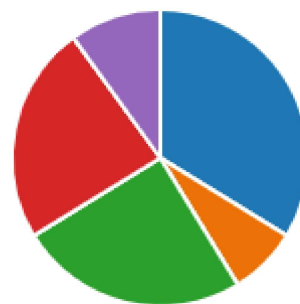
14. Sometimes we see videos of people telling us how good something is so that we might buy it, but what they don't tell you is that they have been paid to say that, and these are really just advertisements that are part of a video that has been made for children on platforms like TikTok, Instagram, and YouTube.

● I think that videos like these sho...	32
● I don't think that videos made f...	41
● I'm not sure	4
● Other	4



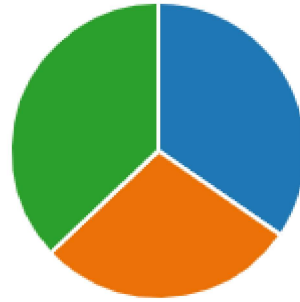
15. Can you tell us how you think you should be able to tell the video sharing platform or service how old you are so that you see videos that are most suitable for you?

● By providing proof with a docu...	27
● By letting them guess my age fr...	6
● By just telling them my age and ...	20
● I don't think I should have to tel...	19
● Other	8



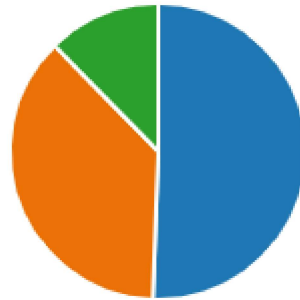
16. Do you know if the videos you are looking at have parental controls?

● Yes	28
● No	23
● I'm not sure	30



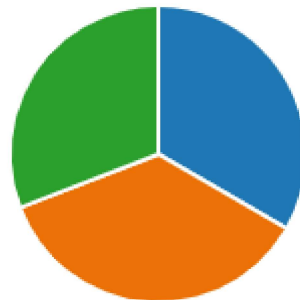
17. Do you think that parental controls should be on all videos that are made for children and young people or should that be up to the parent to put them on?

● Yes, they should be already be t...	41
● No, they should be turned on b...	30
● Not sure	10



18. Did you already know about content feed and how they get the information for that content feed?

● Yes, I already know about conte...	27
● I have heard about content feed...	29
● I did not know anything about c...	25



19. Some children and young people need extra support for reading, writing, hearing difficulties, difficulty seeing or other types of difficulties. If you need extra help for deciding what videos to look at, are there any rules or controls that you would like to see some of the video sharing platforms or companies put in place that might make your online life safer and more enjoyable? Please let us know by adding your comment to the box below.

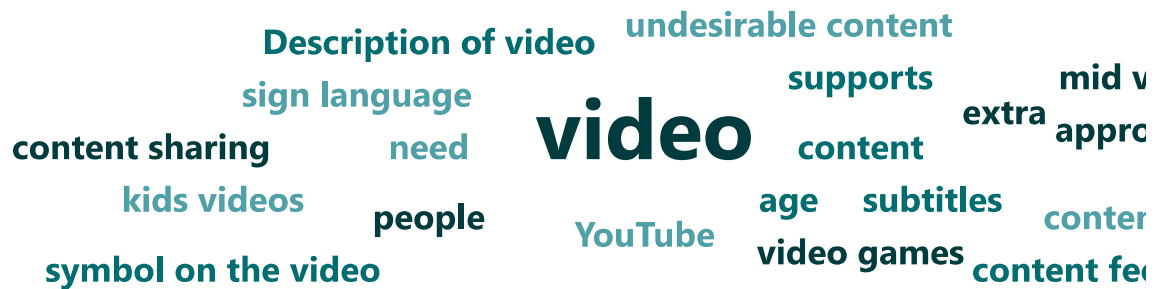
34
Responses

Latest Responses

"I think that there should be automatic subtitles to be turned o...

"Have a section for easy access wirh supports for anyone with e...

11 respondents (32%) answered **video** for this question.



Parent's Comments responses to question 21

1. Platforms should be legally responsible for ensuring that inappropriate context cannot be accessed by users based on their age. It is nonsense to think that parents have the technological mastery to be able to protect their children from harmful content. All adult material should require definitive age verification.
2. Why weren't the dangers of gender ideology and the promotion by radical trans activists of the castration of children listed as one of the dangers lurking online for children? This is more dangerous than the promotion of eating disorders online....
3. You tube has a kids version but it's too childish for my 10 year old. I have blocked explicit content on the adult version but I do worry about something unsuitable coming up. I would like to see a version more suitable for tweens/teens.
4. More awareness and advertising of the full consequences of what children can and are exposed to on social media needs to happen. The true extent of the harm it causes to the vulnerable or immature mind requires much more airtime. I don't believe the lay person fully understands the problems with social media themselves and the above platforms so therefore can't possibly look out for their child using any of the above platforms. France and even the UK seem to be ahead of Ireland in terms of the protection of children on these platforms. Also I believe the platforms themselves and content creators need to have a larger level of responsibility for the content and who can access it. At the end of the day these companies are making money from views as are the creators of the content so they need to be held accountable like any other business. The content creators solely care about views and clicks for advertising and income so who views it or makes the click is of no concern to them as the money is already been made. More awareness needs to be made around how much revenue the above companies make and the content creators and then people would know how valuable their child's time on these platforms is actually worth.
5. Questions that pop up like " what is $9 * 7$ " stops a young and maybe naive person going further with posting
6. Stricter control on what advertisements are played before and during Youtube videos. Several times, when our young daughter was watching cartoons on Youtube, inappropriate ads were played. One was for a horror film.
7. Ban unsuitable content full stop. Why should it be allowed? It's immoral anyway. It's far too much pressure on parents to monitor unsuitable content. The platforms should not allow bad language sexual content or violence.
8. Try not to let underage children on .
9. I agree with banning the use of mobile phone devices for primary school children
10. Make it mandatory that ALL children with & without additional needs do not have access to inappropriate, harmful content
11. I wish there were more towns to follow the example of Greystones, where parents collectively decided they wouldn't let their children have a smartphone before they go to secondary school - and I wish I could live in one of them. Maybe prompting and encouraging such initiatives would be beneficial.
12. Make it safer
13. My children will not use social media or sharing platforms until they are 16. None of their friends do either (aged 10 -13) as their parents are on the same page.
14. Disable online comments preventing online bullying

15. Take off the pop ups and put more safety measures in place for them. Also have safety measures automatically on social media platforms (let adults turn them off) or make it easier for adults to put them on!
16. Children will always find ways around parental controls. The platforms need to be more responsible and face penalties otherwise. Porn can be viewed in harmless seeming apps like pinterest. Tiktok very quickly brings young people through to videos on disordered eating and upsetting emotional content. Youtube kids is a good more though not perfect, but doesn't have enough content for a tween, same as Spotify. They should all restrict content based on age. However, I don't want to submit my child's passport or have their face scanned! I would be happy to set up access through a family account on each platform that allowed more control on access rather than relying on Google family link as some mitigation. The impact on young minds is yet to be seen from apps like tiktok which has such great algorithms that are amazing when properly used but reinforce negative messaging otherwise.
17. Limit the amount of time an ip address can access all of these platforms. Improve age criteria with these platforms my 9 year old set up a Facebook account.
18. "Make it mandatory for age restrictions to be in place & age assurance to used e.g a 10 year old should not have access to social media sites.
19. Encourage primary schools to engage it a mo mobile phone policy, if it becomes the norm no student will feel like they are missing out.
20. Parents have a big responsibility here too "
21. "More online training courses and awareness webinars could be made available to parents to Educate them on the dangers and on how to best protect their children, including how to set parental controls.
22. More tv and radio advertising on the age limits of different services (SnapChat, TikTok, etc) as most parents believe these apps to be for children and to be harmless. When so many classmates who are below the age limit of the services are using them, it is hard to be in the minority of restriction/refusing use.
23. More advertising on TV and radio signposting where parents can go to learn more."
24. More secure for parents to review and unlock before allowing child to unknowingly unlock the programmes.
25. Regulate the platforms taking responsibility for content and sharing.
26. I think video sharing platforms should be made more accountable for what is available on their platforms with large fines imposed on companies that allow unsuitable material on their platforms. They are not in my opinion doing enough. One on my children was exposed to sexual content on another child's phone in a schoolyard despite my child not being allowed a phone and age restrictions I have on devices at home. Blaming parents and making it their responsibility is extremely unfair. I also be in favour of the government legislating for this and banning phones in schools.
27. "Online short free course for parents, a standard, link sent by schools to all school age kids with basic instructions and info on the internet. I am a primary teacher, quite technologically literate. Despite all of this and my rules, my kids have still been frightened by ads for scary movies popping up or ads warning to report abuse if you see it etc. It is infuriating and makes me feel like a bad parent. The dream would be if you can develop an app for irish parents to filter everything basically!! Here is hoping !!
28. Thank you for this survey and all you do. "

29. Not sure
30. "Yes, there should be super strict rules and requirements for online platforms in terms of what is available to be viewed, depending on age.
31. I think the only reliable way to do this is to have mandatory, verifiable authentication in advance of being able to view content on, at least, the popular social media platforms.
32. I also think the current age of 13 should be increased to at least 15 for children to have their own accounts.
33. Children are, in many cases, not emotionally equipped to process much of the unmoderated content they see online and it can have a negative effect on their development and how they perceive real life.
34. Verification should be completed for younger children by a responsible parent/guardian before gaining access to a platform.
35. Would there be an opportunity to create a centralised identification platform that could use federation or a similar tech to log into sites once an initial verification is completed?"
36. Video sharing platforms must be held fully/accountable for the content they create/broadcast. All children are at risk, and additional needs children are even more vulnerable. My main experience is with YouTube Kids. I find the parental controls to be deliberately unhelpful. They will work on one device but can then be over-ridden on another device (Smart TV). The fact a child can simply input the answer to a multiplication problem to over-ride parental controls is a pathetic excuse at child-safety.
37. I think it's shocking that Ireland allows advertising addressed to children in ultraprocessed foods.
38. Make them take responsibility for what they putting up on their platforms. And who it's been aimed at. And should be no advertising to children, food, exercise body
39. There should be laws in place to protect children online
40. Every platform that is providing content to children should have, by law, parental control software built in. All content should have narrative descriptive keywords for parents to quickly read to help decide if the content is okay or not. I use Commonsense Media for a lot of my content information. We have devices in the house which have no parental controls and this causes problems. My children share my Audible and Kindle accounts and have free access to my entire libraries. It requires constant monitoring from me to make sure those devices don't suggest titles to my children that are unsuitable. I wish there were parental control options but there are none.
41. Kids accounts or age accounts should be colour coded, or have a very obvious symbol for parents to know the account is set up correctly, any child under 18 should have parental log ins, once a child has account it's hard to access them. All age accounts should have a similar theme or colour across the different app platforms
42. Moderate all content, such as ads, videos! As I found, there are lots of sexual videos on YouTube, that's why my children don't use YouTube, TikTok, Facebook and Snapchat. I find my children are not safe on those social websites. Thank you
43. Educating them, refresh inform them in school regularly! The reason is that the most of their time they are in school.
44. "Informing parents, including some info in SPHE for the kids.
45. None of the tech company CEOs' children have phones at a young age, which is telling "

46. Yes with parental consent
47. I think the commission needs to engage in an awareness and education campaign. As a parent, its difficult for me to explain in an age appropriate manner the dangers of inappropriate content and excessive use of online tools. If really appreciate some support in these areas.
48. I believe that the control and monitoring is up to the parents however all and any support from the Commission would greatly help to provide structures to online contente.
49. Not sure, but I do know that kids under 16 have multiple accounts, with multiple age's for different reasons...not ideal
50. I think Identification should be used before any account is allowed.to be set up and parental conformation
51. Accessibility for both posting and removing content should be considered.
52. I believe smart devices and social media accounts should be allowable only for those over the age of 18 or under with specific parental consent including identification documents uploaded by parents and that responsibility for the actions of minors accounts should be shared by the parents. I also believe there should be a way of parents acknowledging the use of the accounts, by way of contract or instructional videos which need to be watched before an account is created. Too many parents are clueless of what their children are at or have access to, placing children at huge risk from their own peers, other adults and themselves.
53. Yes platforms could automatically set standard parental controls on under age accounts and accounts where age I unverified. Stopping comments, requests. Messages from unknown people and stopping overage content featuring on their feed. These companies should also be held more responsible for what is on their platforms. When videos or fake accounts etc are reported very little is done about it. Often nothing at all! There should be harsher consequences for people who use the anonymity on these platforms to abuse others especially minors
54. "Influencers on sites like Instagram need to make it clearer when they are using a filter or advertising something. It should be displayed on the video.
55. There should also be more education around online and how what you see isn't real life. "
56. Age appropriate. No advertisements. Confirmation of child's age.Too many children have access to tik tok and posting videos
57. Ban Tiktok
58. Age appropriate content controlled by platform with heavy fines and controls in place by regulators, no advertising to under 18s, parental guidance on platforms and for devices used to access content.
59. "Parental education on danger !
60. Stricter requirements for platforms "
61. Parent education, online safety should form part of special needs overall supports.
62. "The you tube shorts are an absolute disgrace , I've searched everything to be able to block and there is absolutely no way! Yes you can block users of YouTube videos but you can't block the short videos. I think you tube is the worst app ever for children, they could be watching an innocent cartoon and half way through something totally inappropriate pops up. You tube seriously needs to be looked at!
63. The likes of tik tok, Instagram, Facebook under 16s SHOULD NOT BE ON THEM!! I think parents should block these apps for all their children under 16 on their phone. This is not just for the Commission for online safety to safeguard our children, parents need to be on their side too! "

64. Face recognition and age assurance are not an option as they would cause other undesirable effects. Gating is ok but parents need to be knowledgeable empowered and legally responsible for minors. Legislation on minimum age should be clear and enforced at home and in schools so that children do not feel that they are different or at a disadvantage if their parents are more concerned about their welfare and legality.
65. "Parental controls should be on and adds turned off where possible.
66. "
67. Online content should be policed better
68. Only allow it at certain times of the day.
69. There should be an age limit on using them at all
70. Kids shouldn't be able to see much of the content on these sites. Some of the content on kids utube is disturbing. I won't allow my 8 year old to watch anything on her own. Even for my 12 year old I am very concerned that he might go into content that is not suitable. I find it very hard to regulate this. It should be easier to block content based on the child's age.
71. Educate children in school starting at senior infants on how to use the online world and what to watch out for and how. The same way we educate children about crossing the road, strangers that approach them or any other danger in the offline world. If they know what to watch out for and how to behave they can always be safe.
72. Not sure, but anything to provide safety for the kids is valid.
73. The ads in some of the playstore games are not suitable for the age the game is suitable for.
74. Kids you tube is too babyish they won't use it. Then they go to their friends houses or the friends have phones and we have no control over what they see. They watch Mr beast in school. We are not parents anymore we are screen police and it is not healthy for anyone. Kids being exposed to all sorts eg erection ads on daytime TV or they try to stay up a little later over the summer but your sending them to bed for fear they will flick onto someone shopping for their next partner via the appearance of their body parts etc etc etc
75. Stop young kids being able to use tiktok etc. Primary school kids are on it because all their classmates are
76. More dialogue with parents and stricter controls
77. Parental controls assume once an individual is over 18 they don't need any filtering. There should be a way of having a setting for adults with additional needs still being able to have settings on their devices monitored by their parent/carer.
78. "The Commission should have a means to monitor and collate information from parents where issues are not being addressed by platforms. This would provide a means for industry monitoring and feedback to platforms on issues that need to be addressed (to be clear - this should not be a means for escalating issues).
79. In addition, every platform can attract bad actors and I find this survey amusing in respect to having different rules for different categories. The reality is that material that is inappropriate for children will show up in all categories sooner or later. In addition, bad actors will exploit any gaps in monitoring. Platforms need to ensure they are expending the same effort on abuse detection as they are on increasing revenue / viewing hours."
80. Children of all needs are drawn into these sites and have no control of what they will see next these sites easily drawn people down rabbit holes and can end up watching anything with the "up next" lime up is often very random
81. Prevention of harmful material being uploaded/viewed

82. I think that government needs to do more to protect children from harmful content and excessive advertising its frightening how addictive phones are and we don't fully understand the impact they are having on our children
83. Stop allowing people to friend people based on friend suggestions, I was horrified that people my son doesn't know could message him
84. "A parental guide to all ways and uses, safeguarding etc...."
85. I am fairly off with technology but it moves at the speed of light and it very hard to keep up with the changes.
86. With AI becoming more and more relevant we really need to up our game. We have no idea what is going on in the background.
87. It's very scary "
88. Some method on the phone that tells them - "why don't you take a break from your screen for awhile and go get some exercise /talk to someone " especially for boys
89. More regulation and education around social media platforms especially
90. While there's online security courses for parents available through the National Parents Council I think it should be included in the national school curriculum, the way that the stay safe programme is being taught. Now there might be an online security bit in that that I haven't come across yet, if so then the Stay Safe Programme needs to be highlighted more in schools.
91. "More how to for parents and kids
92. Updating information and options
93. There's always a new app or game - ways to keep up with latest trends "
94. I feel that by the time I learnt just how important this is, it was too late for me and my kids. I was of the attitude 'my kids are good and know what's appropriate' or 'they're only watching kids' stuff' (a bit of cartoons on YouTube). However, by the time they had progressed to using TikTok and other platforms it was much harder to then retroactively wrestle devices away from them and to install parental features, device time limits etc.
95. A smart phone ban for under 14's
96. "I selected "age gating" in the previous post because I wouldn't not like my children using AI / camera to "guess" my children's age.
97. I also wouldn't want to be uploading any of their personal data - like a passport to confirm their age.
98. My preferred method would be having a parent, add the child to a "family" account. And allowing parents to decide what age / category suits each individual child.
99. I have social media myself and the videos and posts I have come across on ALL social media is frightening. It doesn't take much to find -violence, gore, sexual, suicidal, hate, bullying and other inappropriate videos, none of which are limited to adults. I have reported numerous videos on Facebook and have had the generic "this has passed out safety standards" reply.
100. X (formerly twitter) has become inundated with horrendous videos of bullying in school, kids fighting and seriously hitting each other.
101. TikTok is full of dangerous "trends" which kids get hooked into watching as they're short clips. I see teens and pre teens who want to be "tiktok famous" and try re-enact these trends which can be very dangerous.
102. Snapchat is another app I dislike, kids able to send hateful photos and videos which disappear. Kids recording themselves doing awful things and saying awful things (bullying) thinking they can't be seen.

103. I think a lot of responsibility is on the parents too. Parents need to understand the dangerous around technology and allowing their children access to technology.
104. More courses in schools for parents would be great, safety nights, email reminders about child safety on the internet etc.
105. "
106. Parental information sessions. As a parent of children with additional needs, I have very little time to navigate the online world and keep up with all the new developments.
107. Phones themselves are causing huge issues for young people and parents on so many levels: they should be banned outright in primary schools and if possible restricted until child turns 15. See Jonathan Haidt's research.
108. "The platform my daughter uses is YouTube kids. She has been told to use this platform only. We have a rule she doesn't go onto YouTube without parental supervisions. My husband and I are not on Social Media so we are not familiar with Tiktok and Instagram. I feel there should be some regulation about the age of person before they can own a mobile device. No matter what controls can be put in place there will always be an individual that can work around this and still be able to access and share content. A national campaign on the recommendation of age before been given a mobile device. Pressure on parents is immense and also you don't want your child to feel left out or excluded.
109. I feel there is a complete lack of awareness on some parents part of the implications of giving your child a mobile device with access to everything "
110. Raising the awareness that regular & ncreased screen time damages mental health. A collective approach (parents & schools) to keep children off phones and screens would be very welcome.
111. Plenty of children are under age watching unsuitable content. Even snap chat needs the Commision for Online Safety to ensure age is real and not just entering a fake birthdate.
112. More control on what shared by the platforms is very important
113. Smart phones are as dangerous as cigarettes in my opinion and we need legislation to make it illegal for children under the age of 16 to own a smartphone.
114. Stricter monitoring of in school use of technology and more robust in school education on safe tech use. Tablets in my child's school were not monitored and children were able to freely download apps and access content that was not school not age appropriate.
115. "There should be no advertising whatsoever to minors online, not only things deemed generally inappropriate but also harmful to the individual or unhealthy, which varies widely from person to person. There is no way to fully monitor the damage so it should not be considered at all, it should all be banned for children.
116. There should be age verification on all content for minors that is age rated in any way above "all ages", & for those under 18 also parental consent. Anything inappropriate for minors should not be accessible to minors in any way at any time. All platforms such as TikTok, Instagram, SnapChat, etc. should require age verification & for those under 18 also parental consent."

117. I work in the safety org at Reddit. So maybe I am not the target demographic. Some of these questions were loaded in one direction or another. I would say that education is the most important thing here. All sites though have methods and tools in place to protect kids. If they don't then regulation should come from government. When people are educated on the tools available they will be more likely to pressure platforms into providing them. I also am not going to let my children have social media accounts until they are 16. This is not to say all content that is not age restricted is not suitable for children but that should be up to the parent to decide. By default all child accounts should be locked down as much as possible and the parent should be forced to removed restrictions as they desire.
118. "Hold the platforms more accountatable.
119. Enforce stronger age controls."
120. Ban these platforms from kids altogether it is the only way control access, my child has additional needs and he is well capable of getting work arounds to parental controls. So my attempts are futile. Snap chat is so risky as parents gave no visibility.
121. I am puzzled that this is about how to use such platforms rather than whether we should let children use them at all. I have answered that I don't implement filters on such platforms because my children don't have any access to such platforms and won't have as long as I can help it.
122. I feel that most of the time creators might not be true to the age restrictions of their content. This could be because they want to drive as much viewings as possible as they will reap benefits from it. That being said there is no true classification of content that will actually stick to the age profile. I've seen this with Youtube Kids where I doubt some of the content has been verified before placed in the platform, as the creators have to put their own classification. So I think the platforms have a big responsibility to accurately classify the content, as the creators do. My suggestion is to use a moderator that can verify the classification and change it accordingly with specific rules... or you can use generative AI to analyse the content and verify that same classification, having some human help on those cases where there can be a doubt.
123. There should be a legal age limit for certain usage and time aloud/limit on each platform.
124. Yes these platforms should provide free internet safety talks in schools
125. In an ideal world, we should not be handing out a super computer to children under 16...at least their brains might be more developed by then
126. Enhance awareness about parental controls AND on how to use them. Share the obstructive, provide some basic training videos share them online make it as easy as possible. I've had problems with youtube kids parental controls and set up so it would be great to have support
127. This needs to be a priority for all company's providing video sharing platforms. So far they have got away with too much and need to be held accountable. They need to enforce stricter age limits on material & any inappropriate content needs to be removed immediately. Twitter has got rid of most of its monitoring staff for this and this is not acceptable. Children are being exposed unnecessarily to inappropriate content and this is going to have a huge impact on them developing into sensible adults.
128. Education on safe usage. Showing stats on how long they spent on it and categories of usage they spent their time on
129. Video Sharing Platforms need to be held more accountable for their content and who it is aimed at.
130. Encourage children to limit phone usage.

131. Limitations to the amount of time they can spend on them.
132. Yes
133. I think if age related controls could be implemented , many of my kids friends had access to tiktok , snap chat at young ages , as girls can easily look older and they all entered false dates of birth. I think a lot of harmful toxic media content , should be age 15 and above and enforcement should be tighter, as despites having parental controls on apps , on qustodian which have blocked my older child from being sent porn . other kids have shown her the images / content on their phones . So even though I am trying to limit / control these on my daughters devices . I have no control over her friends devices and what they show her.
134. Education for children to help with judgement as no matter what control are in place you cannot assume everything harmful will be stopped while you can't fully control what they see on peer devices. Training for parents on parent controls and how to work with their children to monitor usage, discuss content and build good behaviours and judgement around social media given . I think the scope should also bring in AI generated material as it is being incorporated into search engines and productivity products such as MS office tools.
135. A child can enter porn and is exposed to all types on this
136. Unfortunately in todays Irish society children with additional needs are often targeted by bullies using the platforms mentioned above. More needs to be done by the online platforms themselves to prevent this happening. Social media and online platforms need stricter monitoring and controls in place to prevent them from being used by others in this way to cause harm.
137. Clear simple reporting procedures for inappropriate content should be in place accompanied by clear guidance on risks and appropriate controls
138. Ban Social media for under 18s
139. I think that children under the age of 18 shouldn't have access to these platforms and the person using such platforms should be required to submit their passport and verify their account with their finger print or facial recognition.
140. It would be helpful if controls were in place by default. We control the content our kids watch, but even on Netflix content rated U can be inappropriate.
141. The best advice I have come across in relation to this is to watch the online content together with your child as opposed to just throwing them a screen to keep them quiet. Often, there is inappropriate content on seemingly harmless videos such as make up and beauty etc. so even with parental controls in place it is very difficult to keep on top of what your child is viewing online unless you watch it together for a set time i.e. 1 hour a day.
142. My main concern is Snapchat as the messages disappear. My son has only expressed interest in this platform joining secondary school. thank goodness the school have a good policy to mobile use.
143. Bring in legislation banning children from using video sharing platforms and until age where these platforms are least harmful to children
144. I would like to see greater punishments given to content providers who blatantly break the rules. Fines are irrelevant due to the huge incomes they create. A break in the service being provided would offer a greater deterrent.
145. More should be done to control what's posted, even with parental controls on on channels such as YouTube, I have seen videos where people found a way around these controls and included inappropriate content. For example, a kids video showing someone playing minecraft and suddenly the person videoed the phone in their hand and on that phone was an adult video playing. So that happened half way through that minecraft video.

146. I would like to attend training on how to set these parental controls and monitor my children's online behaviour
147. More ability for parents to limit content based on their own expertise e.g. YouTube is the only sharing site that my 10 year old uses but is not allowed their own YT account based on their age(by YT). However because she uses one of our accounts it's hard to control the advertising though I'm monitoring what she watches.
148. Actually after completing the survey, I realised Parents like myself could do with an information session to educate us on how best fo keep our young people safe online.
149. Regular workshops with professionals, journalists, psychologists, high follow influencers. This way children can hear more aspects from different angles. This non-educational rather discussion based sittings suggest common sense choice in children's behaviour.
150. Make the platforms accountable for the content they show the same as traditional media
151. Yea the companies should contribute to child mental health services and child physiotherapist as no matter how much we try, we are losing an entire generation to social media
152. Block advertisement entirely.
153. Parental controls, ability to turn off advertisements for children especially those with sensory difficulties.
154. I have a 14 year old who self verified herself as a 22 year old and while accessing chat channels was exploited online. Despite being a minor and legally not able to provide digital consent in Ireland, internationally none of the platforms I contacted - Reddit, Discord, Twitter or TicTok accepted any responsibility for what happened to her. In their view, none of their 'policies had been broken' due to her self verification. They wouldn't even take down images despite my pleas. It is very hard to balance privacy and freedom of speech with child exploitation, particularly in private chat rooms. At least TicTok have some measures e.g. you can't share images privately. If your child is willing to accept the platforms parental control boundaries, then you have some chance of 'controlling' what they see. But if you have a digitally literate, curious child - you have no chance! The platforms need to put more safety measures in place for adult content - particularly porn which is increasingly violent and denigrating to females (and that's not me being an old school prude). School's SPHE programmes could also do more to counter this online portrayal of sex which is unhelpful for both males and females. Teenagers are learning unhealthy images which then create unhelpful expectations of sex e.g that is ok to choke or be choked. We really are sleep walking into a societal time bomb and it is not surprising that youth mental issues are on the rise.
155. "Expressly forbid devices capable of accessing Internet jn primary schools other than school devices.
156. Funding for annual training in cyber safety for all teachers and pupils from age 9 / 3rd class.
157. Discourage use of phones in 2ndary schools - eg to access curriculum."
158. A module for kids in school and an online module for their parents
159. "There are monitoring subscriptions available however these are not fully usable on iOS due to security. There should be a legitimate option to bypass built in security measures

on iOS so that these third-party subscription monitoring services can allow parents to fully monitor child's internet activity on their iOS device.

160. Also, I feel the 'disappearing messages' format of Snapchat is inherently dangerous and ripe for abuse by bad actors. I believe these chats should be backed up on a Transcript that can be viewed by a parent. "
161. "Recommendations around mobile phones in primary schools (i.e. smart phones not to be used by under 13s)
162. Parents need to take more responsibility for their children's online presence, become more familiar with parental controls etc.
163. Supports should be inclusive for all as standard. "
164. No
165. Moderation of videos should be much improved but also kids should be learning in school and at home about how what you see on these platforms is not real life, it's filtered, edited, advertising, promotion of a person etc.. weekly open discussions in schools in every class at an age appropriate level
166. More education needed in school regarding the dangers of online content. Children find ways of bypassing all the safety features available to access what they want.
167. If a child is uploading a 2nd party should approve before it can be uploaded
168. I think children should be treated equally regardless of additional needs.
169. Its parents responsibility to filter and control what the kids watch nowadays, do our best.
170. Safety of kids first. Default settings should do that. Should not be relying on parental knowledge
171. I don't think fining these organisations works because they are generating such huge amounts of money, I think there needs to be a more effective way to make them responsible for the content.
172. Disappearing messages are a big concern!
173. I believe there should be a national policy for disallowing electronic devices in schools similar to the scheme which was introduced in Co. Waterford lately.
174. It would be very useful if there was a video or other online training for parents on regulating their child's online usage. How-to videos etc. on setting up these parental controls would be helpful. Also coordination of these controls among friend groups would be ideal as my daughter regularly says that she is the only child in her class with online controls, app restrictions etc. I appreciate this would be hard to do.
175. More detailed parental controls which respond to issues that occur on platforms, clear advice to parents on what age platforms are designed for e.g. YouTube, Instagram and Tiktok are designed for 13+, yet 8 year olds have their own accounts. More targeted info campaigns regarding online scams on these platforms for those with additional educational needs online as they are exceptionally vulnerable. As aside but still relevant: Requirement on shops selling devices to help parents set up the device correctly, activate controls etc. and also looking at the pre-installed Apps which are on devices used by children. The influence these platforms can have on younger children buying products online is also an issue. Thank you.
176. "We have a child with additional needs.
177. Better educate parents"
178. Run practical courses for parents - get them to bring their devices into an accessible class and show / demonstrate how to use parental controls. This should be a hand on / practical class.

179. Advise setting PIN codes on adult's profiles. Remind people that you can block certain programs in a child's profile on streaming providers. Advise that kids YouTube is never 100% safe due to the way content creators try to get around restrictions. TikTok should be restricted to 12 and upwards, due to the dangerous "challenges" that often appear.
180. Have tighter restrictions on who is uploading videos and what content is in said videos
181. They could regularly keep a check on the age group that are using these platforms not and regulate all platforms so that the age group cannot go into content that they are not supposed to look into
182. Accessibility is an issue. Kids are more tech savvy than their parents, so safeguards need to be there to assist the parents in safe management of device use & content access.
183. The rise of deep fakes in light of rapid AI developments and the amount of fake news is a concern, especially as I see young people getting most of their information exclusively from online sources. How can we help them differentiate what is real and what is not?
184. When flagging bullying, follow up should be enabled which includes a consult with a therapist, blocking bullies from communicating, informing bullies parents
185. Age limit and parental consent
186. Ensure they can not have accounts under the digital age of consent. Ensure social media providers apply parental control on adding friends on younger children's accounts. Parents vet friend requests of their child so they know who they are talking to online.
187. "Yes. Make educational content (video, presentation, etc.) on the topic of Internet safety and send out this content to parents so that they can discuss it with their children. Or organize a lesson at school (using prepared educational content) on the safe use of the Internet for children. You can also use these two methods at the same time.
188. Thank you for taking care of our children."
189. Dangerous content like abuse videos, bullying etc should be banned and taken down. All content should be vetted before upload to TikTok etc. There is insufficient barriers in place for children even on Kids YouTube they can be exposed to inappropriate content.
190. "Point to reliable and safe sources for key information related to content viewed e.g. HSE.
191. Prompt the child to talk to a safe adult if they are affected by anything they viewed that confused or disturbed them.
192. Clearly state that the content is for people aged over YY and that if the person is younger, the content may be quite confusing or upsetting. "
193. Provide training and information on social media and having an online presence. There is no getting away from social media so why not arm them with the tools and knowledge to use the technology responsibly and get the benefits of it
194. Run workshops in schools for staff and parents/guardians on online safety.
195. not sure
196. We should follow the UK model of adult content only being made available if you specifically request it from your ISP or mobile operator. This is also only available to those over 18.
197. If a young person with additional needs post an unsuitable footage it must be taken down no matter what age they are
198. n/a
199. Parents should enter a pin for any unsuitable content for kids age that's inappropriate

200. Teach them how to use it in school. Teach them like a subject and monitor how they understand it in school.
201. More education aimed at parents.
202. some stakeholder other than the parent / Child needs to limit the amount of "on-line" time children can spend on a device daily. Children don't have the self control to manage this & most spend hugely excessive time on-line. I feel Most parents aren't technology savvy enough to managed (unless they work in the IT field). I feel the result is having a damaging & negative effect particularly on 12 to 16 year olds daily lives.
203. Better monitoring of spam/advertising accounts
204. "I think you have covered it all, however as my daughter goes into 3rd class, I am concerned by the level of bullying that is happening online in chat apps etc. and I believe that this needs to be considered as seriously as the online access to sharing platforms.
205. Many thanks "
206. They could try and enforce a way that the owners of these platforms should require confirmation from parent or responsible adult to prove the child is allowed to use them
207. Enhanced parental controls, different age recommendations
208. Introducing an online platform where users can report safety issues not addressed or incorrectly addressed by the video sharing platforms, for further analysis and action against the platform, if needed.
209. Education. Online literacy, safety and supervised practical experience should be incorporated into all levels of the school curriculum (plus homework exercises involving parental participation). Not just using devices to complete other parts of the curriculum (e.g. maths, reading) but a dedicated 'Digital Life Skills' subject to compliment traditional Home Economics.
210. Access to bad content should be locked and only made accessible by a department person who can verify the person is an Adult and should be done monthly in case of a child breaking into an Adults site.
211. I've no clue on technology. Very basic. & shares me children know more
212. ensure schools technology education is focused on staying safe online, thinking critically and evaluating digital info, social media caution - these to me are more important than using digital info as I think all children now are exposed to technology and being aware they need to be cautious and limit its impact are more important than digital skills for this generation.
213. This is a priority for any parent and social media and digital platforms are causing a massive negative impact on the lives of children and teenagers. I believe this topic should be high on the agenda!
214. I have been using parental controls and selecting content on age rating but still I find too many instances during programmes where content does not correspond to the rating given or content explained at the beginning of the programme. Online platforms don't do enough to identify and filter material based on rating. Children still get exposed to inappropriate material during programmes which have been targeted towards that audience. For example how on earth gun violence, suicide, gory images and nudity with sexual content ok to watch for a 13+ or even 15+child? Still many programmes aimed at 13+ and 15+ show all this in the programmes. When it comes to using copied images and soundtracks, online platforms identify them with their AI algorithms because it affects their revenue but there's blatant lack of responsibility and sheer inaction on their part which is leading to mental and psychological difficulties in our present and future generations. These online giants must be made responsible to do more towards ensuring that young minds get healthy entertainment.

215. Provide a little more assistance
216. If possible educate children about online etiquette.
217. Common parental controls across all platforms, setting parental controls on one device or platform eg google, does not populate it across all platforms and is impossible to gauge how safe any platform is
218. Just a good training for parents regarding parental control
219. "Ban Tik Tok, Snapchat & instagram. It is destroying our children's lives. The content is ridiculous , gives them access to everything everyone and anything,
220. children believe the content is true, & older people are contacting children offering "videos"
221. Of a sexual nature.
222. This country will have a very serious problem in a few years if there isn't something done now to protect our children. All phones with should be banned till they are at least 16, or just have a phone that can ring & text. This is a crisis situation but nobody seems to notice"
223. It's too easy to give a wrong date of birth. Should give ID for Snapchat and gaming
224. I found even with Parental Controls we have come across adult content especially YouTube and Tiktok
225. Stricter rules around children using sites
226. Work with primary schools to discourage students bringing phones into schools
227. More training for parents in how to restrict access to content for kids
228. Normally the parent should be responsible to limit screen time as I do with my child with additional needs, and should be very attentive to what they watch.
229. Not that I am aware of.
230. Completely disagree with question 12(Did not answer). The parent should upload their own documents or verify a child's age. Parents must also hold a responsibility. Why would we want to upload pictures & date of births of our kids to online streaming platforms. Raises a red flag for me.
231. Should have parental consent to access these platforms and that is for all children, not just those with additional needs!
232. Help in educating parents on how to have more control over what they are looking at and educating kids more about online safety and that most of what they are looking at is not real life!
233. Force online platforms to moderate the content published on their respective platforms in line with the age appropriate content guidelines laid out.
234. I think it's too easy for young children to get onto certain apps all they have to do is lie about their age. I have also found that on some children's apps/games that their are people on their messaging inappropriate things. I'm lucky my child told me but not every parent will know the things being said to their kids. He was on Among Us game which I thought was safe enough obviously it was deleted straight away.
235. I'm not sure. I don't believe young kids should have access to these platforms at all.

236. Educational content to warn them of dangers needed
237. More regulation needed from the top down. Parental controls should be more accessible. Sometimes they're almost hidden within the app
238. Nothing comes to mind
239. Ensure schools enforce rules and discuss them with parents and children

240. Enable voice over messages sonuser is aware of content age the content is aimed towards
241. No
242. Work with Media Literacy Ireland
243. keep photo thumbnails when search results come up as alot of special needs relate to front pictures of the video or song they are looking for. Also it would be good to have an option that says click here to "skip add" with an arrow as sometimes my daughter can't work out when to press to skip the add as some you have to watch the add to the end and some you can skip after 20 seconds
244. Provide child-friendly videos that educate children on the value and dangers of online usage. Children need to see and hear a voice other than the parents. Perhaps have small infomercials before or after the news or children's programming on television.
245. Not sure
246. Make sure reported content is handled promptly and correctly, giving feedback to the person who reported these.
247. "The video platforms have to be positive and safe for any online users.
248. Children videos must have safety control on Ads and misleading information about promising things that can't happen in the real world.
249. "
250. Passport should be used to know the age of the child and only information for the child should be send to the mobile.
251. Yes - provide free advertising for childline and associated children's charities and helplines
252. Ban smartphones from schools
253. More laws should be brought in to protect innocent children. Their young minds are unable to process so much content. This can be harmful.
254. "Education and informative ads about parental controls and inappropriate content should be placed into those online places where they can be seen and targeted i.e. if you've watched 1 hour straight on youtube kids or roblox an advert should interrupt asking them to show it to their adult before proceeding.....
255. also short tv adverts on telly eg at news time or during coronation street, as lots of grandparents have laptops and tablets and allow kids to use them without understanding the risks. They certainly wouldn't have the tech savvy to start setting up different accounts and different controls etc"
256. Online Parent education
257. Maybe courses for parents about how to use Parental Controls
258. While I appreciate the intent of age checking I do not want to provide id and birthdays to strangers online. I much prefer other controls. Unfortunately it is mostly the responsibility of the parent to ensure kids aren't able to access inappropriate content, and information on how to do that should be more widely available.
259. Age assurance should be required for social media apps like Facebook, Twitter and Instagram. Greater education in schools needs to take place and greater education for parents to teach their kids about online behaviour. Sellers of digital devices should be mandated to provide parental controls on all devices.
260. Videos etc for Tiktok, YouTube that starts with appropriate content with very inappropriate content within to fool the parent controls. More moderating please
261. Be more vigilant in ensuring that companies remove harmful content

262. Ban phones from schools, require phone manufactures to provide age appropriate operating systems for mobile devices
263. Age verification using government identification for all Social Media and messaging apps with a strict user age of 14 years
264. "
265. We don't let our child on tik. Tok YouTube YouTube kids our Instagram. Full of total junk and simply not safe to not be sitting there with them.
266. Suggested age gating as tbh no company will monitor passport upload etc so in reality can't see it working. Doesn't mean I think it's right.
267. <https://culturereframed.org/> should be shared with everyone. Useful clear guidance to stay informed
268. Good luck
269. "
270. I think both the content creators and platform should be responsible and accountable for the content created and shared. There should be strict laws in place to enforce the rules.
271. Vigorously promote the widespread implementation of the Greystones initiative in primary schools. No child should have unlimited access to the internet / social media etc. etc. etc.
272. I'm not sure
273. Ban them !
274. Only allow them to be available for 18+ and putting the responsibility on electronic device companies, media platforms and those who make the posts etc
275. Produce one trustworthy document that parents can access easily, to explain how to monitor content and set limitations on each of the main platforms. Allow comments on same so parents can provide each other with additional information.
276. There needs to be education given to all parents and it should be mandatory. The school I am in provided Internet safety talks. It was brilliant but I know from attending that the parents whose children have phones were not there. Children in my child's class are using certain apps which are dangerous and not suitable for any child
277. All apps should have a "Grown Up" mode that an adult has to enable so children don't get exposed to content inappropriate for their age
278. I think they should all be banned imo.
279. Better monitoring of online bullying
280. Smart phones should have an age limit 13+
281. Additional advisory warning. Allow and act on feedback from parents.
282. Making sure there are no gaps or missed content that are not age appropriate
283. The age limit should be higher and only accessible through more strict age verification process
284. Legislation needs to be put in place to enforce policy rather than it being discretionary
285. It is my opinion that no child under the age of 16 should have access to social media. There is absolutely no evidence to support it being beneficial in any way whatsoever. Social media accounts should be linked to official ID documents e.g passport or driver's licence. You can only set up a social media account by providing verification of either document. Children under 16 should not be allowed to have social media accounts.
286. "1. The age of digital consent in Ireland should be increased to 16

287. 2. Classes for parents as none of these things are failsafe and parents need to be more vigilant/discuss with their child more what they might see etc.
288. 3. More on the dangers... Particularly around online bullying leading to suicide.
Online eating websites leading to body dysmorphia and disordered eating
289. Consent and tell not to video share, even with trusted people.
290. Educate parents in simple lay man's terms through the schools network do parents realise we are all in the same boat trying to protect our children
291. I wish all phones were banned for use in secondary schools. They have no purpose on school grounds. My daughter is asked to use for Google classroom and says they all just go on snap chat. I wish smartphones were illegal for under 18's.
292. they should be able to restrict these platforms to age appropriate. So if you are 6 that you don't have tiktok. Snapchat should be banned its a purely bullying platform. they can group chat on whatsapp if required. you tube should be age controlled.
293. As far as I understand inappropriate videos at the moment have to be reported by many viewers before they are taken down; I've age restriction on my kids YouTube yet I don't seem to have access to what they've watched; on Netflix kids account seems to allow stuff that's not age appropriate; Disney channel has had wrong age ratings for movies - weren't actually kids movies and showed "U"; movies like Home Alone and many more older movies have violence in that I find inappropriate for young audience or even for myself; there used to be lots of fake Peppa Pig videos on Kids YouTube, and many videos with hurtful underlying/built in/hidden messages - I'd love to hear that there's a way to eliminate those videos; at the moment, when I want to block a channel on kids YouTube, I need to start the video, to get "block the channel" option= there should be an option to block these channels without watching any videos, and to block channels with specific area like video games etc that don't suit some families; I've YouTube premium for the whole family= no ads, therefore I can't comment much about ads, I just know some Sky apps like Ninja kids show ads, they seem ok so far. Myself and my kids have stopped watching YouTube many times. Unfortunately on TV there is no option to delete Kids YouTube app. I understand sometimes it's useful, even school recommends some alphabet stuff. Thank you for caring.
294. No comment as I'm not parent of child with additional needs
295. maybe a way to link accounts privately so any harmful content would be flagged
296. Yes Tiktok is totally unregulated and I would argue does not have a sufficient process of verification of account creation- I currently have an open case with data protection with them whereby they enabled a fake account to be created using another person's data to purport to be a young girl. I am frustrated with the fact that providers self regulate and parents are oblivious to the fact that children are essentially using self censorship. It is a slow process to get any accountability and regulation is too slow and unfortunately not implemented including GDPR- parents need to be educated as to the dangers of children and adults giving away their data unintentionally or data theft including personal identification data I also know a friend whose facebook account was hacked purporting to be another person and they also cannot get any accountability - it is very hard to get any support. I also object to the wording of one of the questions whereby the answer may be skewed in favour of platforms indicating data is tailored to individuals - there should be more options or a free comment text answer for this as it only allows answer stating they are not aware or are aware as such agree with the comment.
297. Promote no-phone policies in schools
298. Visits to schools to provide info sessions on these platforms/ simple visuals or video clips to explain

299. "Provide classes to teach parents exactly how to put on parental controls. would be useful.
300. Getting a law to ban phones for primary school children and ensure they are not used during school hours in secondary school. They are harmful, stricter rules can be applied but it needs to be same rule for all"
301. "1) our culture, and state policies should affirm that it is the parents' right and responsibility to screen content according to their wishes, for their own children - and options (technological and social) should be promoted to facilitate that parental choice;
302. 2) it should not be the first option - by state and society - to mandate that everybody else and every sphere of online activity has to bend towards being a functional babysitter for everyone's children - and turn into a virtual panopticon against adults and children both."
303. I'm very pro the banning of mobile phones for primary school age kids. If we all do it then nobody is left out. It is just how it is. Really like what they did in south dublin (I believe). Can we make it national?
304. Advocacy issues
305. "Best is for children to not have access to tablets, phone etc ... and banned social media platform (tik tok, insta....) as they don't bring any values to education or development.
306. "
307. Creation of an online platform aimed specifically at primary school age children where the parent must approve/deny posting of pics, posts, etc on a live basis. That way if inappropriate content is being produced, an adult is responsible. The platform should also provide mediation of some sort for any disputes which arise where the parents/guardians can discuss the post.
308. Rather than trying to police the use, which children can figure out how to bypass, inform children about the motives behind the tech industry, that using video sharing platforms is free because THEY, the children, the users, are the product. Some platforms can be very creative with video editing features, children engaging with these features might help encourage children to be more active in their use (creating) rather than passively scrolling.
309. Have the ability to link these platforms to a parents mobile/tablet so that the parents at all times have the power to oversee exactly what the children are watching and getting involved in
310. The commission should advocate strongly for the government to introduce strict controls on content for platforms. I think they should be regulated in similar ways to broadcast media
311. Phones are the main issue for me, kids have access so easily. We haven't given our 12 year old daughter tik tok yet but she still gets sent stuff from the platform from her friends. Also, use app which we pay for to control phones but the apps will cover something and often cause issues with the phones.Very frustrating.
312. More information needed on parental controls for different gadgets such as laptops vs phones etc and how to link these as some apps may only work in one setting.
313. More education to parents on the true harm of these platforms to help them steer their children away from them for as long as possible and co-use them thereafter. Even adults struggle to monitor their usage habits - underdeveloped child brains haven't a hope against the algorithms.
314. "1. They could make it compulsory for parents to attend a 'how to keep your kids safe online' course.

315. 2. They could fight to make it illegal for children under a certain age to use a smart device.
316. 3. They could fight for child friendly phones be invented."
317. I wish none of it existed for my children. Business benefits but no person benefits for kids
318. Work harder to keep children away from social media until they are in secondary school. Provide more training and resources for teachers and parents. Have more advertisements related to the potential harmful affects of all of above.
319. Create an educational environment where those platforms are eradicated as they have no place in the classroom
320. "Fines for the platforms who are not enforcing age restrictions.
321. Platforms should be monitored by external bodies which are government funded and should have the ability to issue large fines where harmful content has not been handled in effective or appropriate ways."
322. Perhaps in-school talks to kids making them aware of the dangers of online platforms, misinformation and real life examples of harm/problems caused by regular use of platforms. Include age appropriate talks on the addictive nature of video games, pornography and the truth behind the pornography industry and the harm viewing this, violent games and over sharing by children (photos and videos used to ridicule etc) can cause using real life examples.
323. "Fine the platforms if content isn't taken down after being reported.
324. "
325. It should be law that kids under 16 have no data/internet on their mobile phones. Smart TVs are also an issue going forward as they all have streaming platforms.
326. A default Automatic 'time out' of social media platforms after 1-2 hours, unless settings are changed
327. "if we are going to take on line safety seriously, we understand the risks, the devices on which underage children can access such content should be better regulated... if you have to be over 18 for social media, there needs to be much more robust regulation and penalty where underage use is identified and parents need to play their part in this.
328. Our children need to be given the opportunity to be innocent, and also space and time to develop their own views, not force fed constantly by on line vultures.
329. Technological advances were intended for the adult working world to provide efficiencies, NOT to e a toxin for our youth. "
330. Stricter controls on targeted advertising
331. Bring in new Laws so the makers of harmful content are prosecuted.
332. Educate children and adults. Ongoing information at school in the community.
333. I think it needs to be done as soon as it is possible for children
334. Have direct connection/ regular exchange with the platforms to ensure their collaboration on the matters. Most of them are very interested to get this right for the right audience
335. There should be a state platform, an RTE version of YouTube. No likes or comments.
336. We shouldnt allow phones in the primary school .
337. Support the parents by providing workshops to explore issues
338. Why can't we do what France did and put a legal age limit of 16 on mobile phones/tablets. So kids can't own one legally u till then. Then I think there should be

consequences for publishers of inappropriate content not labelled/restricted from children in these platforms

339. So scared of what internet can do to my little ones....
340. Ensure that content can be taken down when it causes hurt to someone.
341. I believe protection for children from harmful content begins with how the smart device is set up by the parent to tailor the restrictions to their age. Providing how to guides to parents would be the best support for parents.
342. stop watching you tube, tik tok Instagram children under 15 of age

Children's comments responses to question 19

1. Some children and young people need extra support for reading, writing, hearing difficulties, difficulty seeing or other types of difficulties. If you need extra help for deciding what videos to ...
2. Should be able to email or notify the companies if you have any of these difficulties
3. No comment.
4. "No
5. "
6. unsure
7. no
8. Add subtitles and sign language. Audio Description of video for blind people
9. Yes, please put more clear and simplified rules for content sharing, links and how paid advertisements mid videos work.
10. Toodloo
11. I don't need extra help
12. Auudio description so they can hear what it's about not just read it.
13. idk what ur on my moms making me do this
14. Sign language writing it out what they are saying in brades
15. I'd like to press a button for it to be read out if I can't read it
16. Yes
17. I think if you have trouble reading that you would not be on those apps. But there are some supports for that. Also the apps are catering for the majority.
18. Im not sure
19. I think that you should be able to have a filter that shows someone translating videos into sign language for people with hearing difficulties.
20. Let me watch what I want
21. Small description to let you know the content
22. Subtitles
23. When making the video put the oldest age that can watch it.
24. All videos should be checked by a safety person from the platform before they can be approved to be uploaded online
25. Have a section where you can highlight need for extra support, where you can choose what you require eg. subtitles maybe.
26. Easier way to remove undesirable content

27. Maybe have boxes come up on screen before videos to suggest getting parental advice before playing the video, like a second opinion
28. yes
29. Sometime language is not proper according to age group so need improvement in this case
30. A setting to say edit what they think you to watch or edit the content field with tags or ban certain tags or disable content feed as an option and YouTube should add back the dislike counter so that I can see how good the videos are and also the ability to rate ads with a 5 star system.
31. Additional parental controls. Tickbox to say the user has Additional needs
32. I would only like to watch video games because video games are just made up games on PlayStation, Nintendo or Xbox.
33. Age appropriate content; that everything is checked before it's posted on YouTube; clear and longer description on YouTube kids videos.
34. The video should be verified by a teacher and have a symbol on the video so that they know it will actually be able to help them
35. Have a section for easy access with supports for anyone with extra needs so they don't have to scroll through all the videos
36. I think that there should be automatic subtitles to be turned on for deaf people or people with worse hearing

IWF response to Coimisiún na Meán's call for inputs, online safety:

Developing Ireland's first binding Online Safety Code for Video Sharing Platforms

About the Internet Watch Foundation:

The IWF is a UK based charity that works in partnership with the internet industry, law enforcement and government to remove from the internet (with the co-operation of industry) child sexual abuse images and videos wherever they are hosted in the world and non-photographic images of child sexual abuse hosted in the UK.

- We exist for public benefit and perform two unique functions in the UK: We provide a secure and anonymous place for the public to report suspected online child sexual abuse images and videos, and Non-Photographic Images include cartoons, drawings, computer generated imagery (CGI) and other non-photographic depictions of child sexual abuse that are deemed to have breached sections 62-69 of the Coroners and Justice Act (2009).
- We use the latest technology to search the global internet proactively for child sexual abuse images and videos, then work with partners to get them removed.

In addition, the IWF has established reporting portals – places to anonymously and safely report online child sexual abuse imagery – in 49 countries around the world, serving 2.5 billion people.

There is a Memorandum of Understanding between the National Police Chiefs' Council (NPCC) and Crown Prosecution Service (CPS) that governs our operations. This ensures immunity from prosecution for our analysts and recognises our role as the “appropriate authority” for the issuing of Takedown Notices in the UK. Operationally, we are independent of UK Government and law enforcement but work closely with both.

The IWF also plays a vital role in providing the internet industry with several quality-assured technical services to prevent the spread of known child sexual abuse images and videos online and to stop the upload of images in the first place. These include image hashing utilising Microsoft's PhotoDNA, a URL blocking list of live webpages, keywords list, domain alerts, payment brand alerts, newsgroup alerts and simultaneous alerts (for US companies only). Key to this is our trusted relationship with the internet industry which enables us to act as a broker between them, and government and law enforcement.

Our work is funded almost entirely by the internet industry: 90% of our funding comes from our members with the remaining 10% of our funding coming from the .UK domain name registry provider, Nominet who fund our work as one third of the UK's Safer Internet Centre.

The IWF has previously received additional Government funding for specific projects and is open to further diversifying its funding mix in the future.

We are a charity registered in England & Wales with an 11-person Board of Trustees of which, eight are independent members and three are industry representatives. The IWF Hotline is audited biennially by an independent team, led by a family court judge, and the report published in full.

Question 1: What do you think our main priorities and objectives should be in the first binding Online Safety Code for VSPS? What are the main online harms you would like to see it address and why?

The Internet Watch Foundation's remit is outlined in the background information at the start of this submission and therefore our response to this consultation is focused on ensuring that the Online Safety Code addresses the issue of child sexual abuse and exploitation online.

We believe that this is important, not only because it is one of the most egregious harms online, but also because there is clear legal certainty over what is and isn't illegal and complements and works with other existing legislation. The Digital Services Act (DSA) provisions have recently come into effect which will require platforms to take a much more proactive approach to addressing harms on their platforms. The DSA will require platforms to assess the level of risk their services could be abused by bad actors and requires them to take steps to address these risks. The DSA complements legal requirements already in place through the e-commerce directive, which also requires companies to "expeditiously" remove illegal content once they become aware of it on their platforms. This could be either through their own teams of engineers and moderators discovering it, the public reporting it or trusted flagger programmes or organisations like hotlines and helplines bringing this content to their attention.

We also believe that it is important to tackle this type of content because mechanisms already exist to prevent the upload and spread of this imagery. The IWF provides technical services to its membership which helps them keep their platforms free from known child sexual abuse material. This includes image hashing technology, webpage blocking, and keywords, being the three main services that would be most applicable to video sharing platforms.

The IWF also has an interest in ensuring that children cannot access content that is age inappropriate for them and we are particularly concerned about children's free and easy access to online pornography. We are keen to see the application of age verification, assurance and estimation techniques on video sharing platforms that is appropriate to the level of risk that they pose to children. For example, a video sharing platform that is focussed solely on the distribution of adult pornographic content should be ensuring that it is taking steps to verify that users accessing their services are over the age of 18. For other sites not offering such content and where the risk is lower, it may be more appropriate and proportionate to use age assurance or estimation technologies.

We are also keen to ensure that in relation to adult pornography, that platforms are also age verifying and obtaining the consent of the people appearing in the images and videos that are uploaded to the platform. This will help to stem the stream of new child sexual abuse images potentially uploaded to adult websites and will also, hopefully, reduce incidents of intimate image abuse on these websites too.

In summary, these are also all areas that are required to be covered in the transposition of the Audio-Visual Media Services Directive (AVMSD) and, can, as the consultation points out be developed as part of the Online Safety and Media Regulation Act 2022, of which we are of course supportive.

Question 2: What types of online harms do you think should attract the most stringent risk mitigation measures by VSPs? How could we evaluate the impact of different types of harms e.g., severity, speed at which harm may be caused? Is there a way of classifying harmful content that you consider it would be useful for us to use?

As set out in response to question 1, we are keen to ensure that the most egregious harms on the internet receive the greatest level of attention and focus from regulators. We are most concerned to ensure that illegal content and specifically, child sexual abuse is covered.

Question 3: Do you have reports, academic studies or other relevant independent research that would support your views? If you do, please share them with us with links to relevant reports, studies or research.

The Internet Watch Foundation's annual report for 2022¹ details information which may be relevant and useful to reference as an evidence base of why greater controls are needed online. In terms of headline statistics in 2022, we assessed 375,230 reports of suspected child sexual abuse material and confirmed 255,588 reports as containing illegal content.

In the last two years, we have seen a doubling in the most severe forms of child sexual abuse, as we confirmed in 2022, 51,369 reports of Category A child sexual abuse material up from 25,050 in 2018.

We are also extremely concerned by the rise in self-generated child sexual abuse content. This is where children have been groomed, coerced, tricked, or deceived into producing images and videos of themselves and have then shared them online. In 2022, we removed 199,363 reports containing self-generated child sexual abuse material and this now accounts for three quarters of all the content we have actioned for removal. The 11-13 age range remains the fastest growing age range appearing in this content, but in the past year we have seen a 60% increase in 7-10-year-olds appearing in this content.

The IWF has also been responding to Ofcom's preparations for the Online Safety Bill in the UK, you can read a copy of our submission to their call for evidence on our website², which may also be useful in helping to further shape the response in Ireland.

Another useful report that you may want to consider, was published by the Australian e-safety commissioner in December 2022, which was the first regulatory report anywhere in the world which provided insight into how the companies (Meta, WhatsApp, Google, Microsoft, Skype, Omegle and Snap) responded to the first regulatory notices for CSE/A as part of the Basic Online Safety Expectations determinations 2022.³

Similarly, Ofcom has also produced its first report on Video Sharing Platforms⁴ which provides further insights into some of the measures we have outlined in our response to question 1.

Question 4: What approach do you think we should take to the level of detail in the Code? What role could non-binding guidance play in supplementing the Code?

The IWF has always advocated for a principles-based approach to regulation and urged Government and regulators not to be overly prescriptive in their approach to regulation. We

¹ <https://annualreport2022.iwf.org.uk/#>

² <https://www.iwf.org.uk/media/tnelu2yi/online-safety-cfe-response-form.pdf>

³ <https://www.esafety.gov.au/sites/default/files/2022-12/BOSE%20transparency%20report%20Dec%202022.pdf>

⁴ <https://www.ofcom.org.uk/online-safety/information-for-industry/vsp-regulation/first-year-report>

believe that primary legislation should set the framework of what is expected of those in scope of regulation which gives regulators the ability to be flexible in their response in order that regulation continues to keep pace with changes in technology.

We believe that differing harms may require different legislative and regulatory responses. It is also important to note that many of these platforms are unique in the way they are designed and no one platform is established in the same way another is, despite the fact they may appear to have very similar characteristics.

It is important that the regulator regime is also flexible enough to work in partnership with other regulatory regimes, such as for example, the regimes on data protection.

In terms of the options set out in the consultation document, we would favour a mixed approach in the code (Option 3). On some issues like CSE/A we would like to see some element of prescriptiveness for example in setting out some of the options that a platform can take to prevent CSAM from appearing on their platforms, such as utilising the tools and services the IWF has to offer in respect of Image Hashing, URL blocking and Keywords. But not all these services may be applicable to a video sharing platform, and they should of course be given the flexibility to prove if they are not deploying these measures that they are mitigating the harm in some other way to the same or preferably improved standard.

Question 5: What do you think would be the most effective structure for the Code? What are the most important factors we should consider when we decide how to structure the Code?

We believe that the greatest focus for the Code should be on the areas of harm that cause the most damage to society. As set out at the start of this submission our interest is to ensure that tackling the spread of child sexual abuse is a priority in this Code.

Tackling illegal content must be a priority, but platforms will not get their approach to this right unless they are able to assess the level of risk, ensure they have effective terms and conditions and are enforcing them and are also taking steps to moderate content. We would urge a holistic systems and processes-based approach in the development of the Code.

Question 6: How should we design the Code to minimise the potential for conflict and maximise the potential for synergies in how platforms comply with it and the DSA?

We believe that regulation should be flexible enough to operate with other regulatory regimes such as the Digital Services Act. In respect of the provisions in the DSA and referred to within this consultation document, we expect that as a hotline, if we were notify a platform in scope of the proposed regulation, that they would act on a notice and act expeditiously, in line with the terms set out in the e-commerce directive to remove the offending illegal content.

Secondly, the DSA sets out steps platforms must take to risk assess the likelihood that their services could be abused to host or facilitate illegal activity. We believe it would make sense if the risk assessment criteria in the DSA are aligned with the provisions within this Code to ensure that companies are not having to carry out multiple risk assessments which could be confusing, burdensome, and risks a lack of alignment between regimes, with one regime telling them they have to do something one way and another regime being in direct conflict, which must be avoided.

We do, however, agree as we set out in response to Question 1 that this Code does represent an opportunity to build on the DSA provisions by adding additional obligations in areas such as

age verification or on directing platforms on the tools and services they should be using to prevent the spread of illegal content on their platforms.

Question 7: To what extent, if at all, should the Code require VSPS providers to take measures to address content connected to video content?

The IWF is supportive of the application of the non-exhaustive list of 10 measures that need to be taken by video sharing platforms to comply with Article 28b of the Directive.

Question 8: How should we ask VSPS providers to introduce a feature that allows users to declare when videos contain advertising or other type of commercial communications? Should the Code include specific requirements about the form in which the declaration should take? What current examples are there that you regard as best practice?

The IWF does not have a view or anything to add in response to this question.

Question 9: How should we ask VSPS providers to introduce and design a flagging mechanism in the Code? How can we ensure that VSPS providers introduce the mechanism in a user-friendly and transparent way? How should we ask VSP Providers to report the decisions they've made on content after it has been flagged? To what extent should we align the Code with similar provisions on flagging in the DSA?

As in answers to previous questions, regulatory alignment with other regimes is important and therefore we favour alignment with Article 16 of the Digital Services Act which sets out criteria for trusted flagger programmes in respect of illegal content. As a hotline providing notice and takedown, we would expect that this provision would cover the IWF and other hotlines. Of course, in Ireland there is the Irish Hotline, and we would also anticipate that they would be a "trusted flagger" of content to video sharing platforms.

We believe that reporting process in place on platforms should be clear, easily accessible to users and clearly set out in comprehensible terms and conditions platforms have in place.

Question 10: What requirements should the Code include about age verification and age assurance? What sort of content should be shown by default to users who are logged out or in private browsing mode and whose age cannot be verified or assured? What evidence is there about the effectiveness of age estimation techniques? What current practices do you regard as best practice? Where accounts are not age verified should default privacy settings be used, should content default to universal content and should contact by others be more limited?

As stated elsewhere in our response to this consultation, the IWF is supportive of the introduction of age verification, assurance, and estimation procedures and that this should be proportionate to the level of risk a platform poses dependent on the content it provides.

This, however, does not sit within the IWF's area of expertise and we therefore don't feel best place to add anything to our previous answers on this question.

Question 11: What requirements should the Code have in relation to content rating? What do you consider to be current best practice? What experiences have you had using content rating systems on platforms and do you think they have been effective? What steps could we ask VSPS to take to ensure content is rated accurately by users?

The IWF is not best placed to respond to this question, but as the consultation sets out, the standards suggested through the Irish Film Classification Office sounds like a good possible standard to align to. We have a good relationship with the British Board of Film Classification (BBFC) and suggest there could be some alignment of approaches between the two organisations.

Question 12: What requirements should the Code have in relation to parental control features? How can we ensure that VSPS providers introduce the mechanism in a user-friendly and transparent way? Can you point to any existing example of best practice in this area? Should parental controls be ‘turned-on’ by default for accounts of minors or where age is not verified?

The IWF supports the active involvement and interest of parents, guardians, and carers in keeping their children safe online. We believe that they should have access to tools and features that enable them to protect their children online.

We have seen through the introduction of the Age-Appropriate Design Code several examples of best practice from platforms, in terms of ensuring Childrens' accounts are private by default, that children cannot be discovered by adults as part of their friends' suggestions and some companies have also introduced measures which set-up sleep reminders and limit screen time provisions for children.

It is important that both children and their parents, guardians and carers are aware of the availability of these tools and products, they are easily accessible and available to users and easy to set-up. We support many of the suggestions made above, such as ensuring that childrens' accounts are set to private by default.

Question 13: What requirements should the Code contain to ensure that VSPS provide for effective media literacy measures and tools?

The IWF is supportive of media literacy measures and tools as part of the Code.

Question 14: How should we ask VSPS providers to address online harms in their terms and conditions in the Code, including the harms addressed under Article 28b? How should key aspects of terms and conditions be brought to users' attention? What examples are there of best practice in relation to terms and conditions including content moderation policies and guidelines?

The IWF believes that platforms should be highlighting to users that illegal content such as the distribution of child sexual abuse material is not tolerated on its platforms and that should such be content be discovered on a user's account, they will have their account immediately suspended and the content referred to the relevant law enforcement agencies.

Question 15: How should we ask VSPS providers to address content moderation in the Code? Are there any current practices which you consider to be best practice? How should we address automated content detection and moderation in the Code?

Companies should all have procedures in place to detect and prevent the distribution of child sexual abuse material at the point of upload. The IWF offers tools products and services which assist video sharing platforms in complying with this, by offering image hash lists, webpage blocking and keyword terms as the most appropriate and applicable services to video sharing platforms. Much of this can be automated by companies, to automatically report 100% matches against this hash list and if companies are deploying PhotoDNA there are tolerance

levels, they can set to detect similar content, where one or several parts of an image may have been altered to avoid detection processes.

As highlighted in response to other questions, video sharing platforms

Question 16: What requirements should the Code include about procedures for complaint-handling and resolution, including out-of-court redress or alternative-dispute resolution processes? To what extent should these requirements align with similar requirements in the DSA? What current practices could be regarded as best practice? How frequently should VSPS providers be obliged to report to the Commission on their complaint handling systems and what should those reports contain? Should there be a maximum time-period for VSPS providers to handle user complaints and if so, what should that period be?

The IWF does not have anything to add in response to this question.

Question 17: What approach do you think the Code should take to ensuring that the safety measures we ask VSPS providers to take are accessible to people with disabilities?

Terms and conditions, user safety functionality should be easily comprehensible to all users. Best practice in this area could include easy read versions of terms and conditions. Other than this, the IWF is not best placed to respond to this question, other than to say that it is important that people with disabilities are given support with their online experiences.

Question 18: What approach do you think the Code should take to risk assessments and safety by design? Are there any examples you can point us towards which you consider to be best practice?

The IWF supports this regulation's focus on systems and processes platforms have in place to protect their users and encourages them to take a safety by design approach, based on a risk assessment process which is conducted by both the company and regulator. Whilst we are supportive of the provisions in the Digital Services Act that focusses on Very Large Online Platforms (VLOPS), it is important to consider that very small, fast-growing platforms may also be at risk of causing high harms for users. It is important that there is good engagement within start-up communities of their regulatory obligations and ensuring that they are supported both in their desire to grow but do it in a way that is safe and secure by design.

It is important to also consider that future EU regulation related to preventing and combatting child sexual abuse is also based on a platform's ability to assess risk and respond accordingly to the threat and risk that they pose.

Another regulatory approach which is being taken in the UK includes the introduction of a duty of care on platform providers to ensure they are keeping users safe on their platforms. In Australia, these take the form of Basic Online Safety Expectations (BOSE).

Question 19: How do you think that cooperation with other regulators and bodies can help us to implement the Code for VSPS?

We agree that collaboration with other regulatory bodies, such as those outlined in the consultation will be important and we actively encourage the involvement of relationships with the global regulators network and ERGA. We also would encourage you to develop relationships with service providers such as the IWF of datasets which can help keep video

sharing platforms free from the spread and proliferation of child sexual abuse material on their platforms. It would be particularly beneficial to recommend the adoption of these services within the code.

Question 20: What approach do you think we should take in the Code to address feeds which cause harm because of the aggregate impact of the content they provide access to? Are there current practices which you consider to be best practice in this regard?

The IWF has nothing to add in response to this question.

Question 21: Do you have any views on how requirements for commercial content arranged by a VSPS provider itself should be reflected in the Code?

The IWF has nothing to add in response to this question.

Question 22: What compliance monitoring and reporting arrangements should we include in the Code?

The IWF would be happy to consider assisting with compliance monitoring and reporting arrangements, by providing data that we have on the extent of harm on platforms, our annual report already provides detailed information on most of this and could have a role in helping to evidence the prevalence of child sexual abuse online.

We would, however, draw the line at wanting to be involved in any enforcement action directed towards companies.

It would be beneficial if monitoring and compliance arrangements were able to include information from providers about the amount of attempts or “hits” against IWF services such as image hash lists, webpage blocking lists would be helpful in us further understanding the prevalence and how effective these services are at preventing viewing offences or the upload and further distribution of this illegal imagery.

Question 23: Should the Code have a transition period or transition periods for specific issues? Which areas touched on in this Call for Inputs may VSPS providers require time to transition the most? What time frame would be reasonable for a transition period?

It is important that companies are given sufficient time to prepare for regulation, however, the regulation of video sharing platforms through the EU’s Audio-Visual Media Services Directive, should have already commenced. It could be reasonable to suggest that companies could already be taking steps to protect their users based on best practice from other jurisdictions.

We would urge a swift adoption of the Code but do recognise that companies may need sufficient time to prepare before the enforcement aspects of the regulation take effect. As we have seen with the development of the Digital Services Act, the enforcement aspects of the regulation have taken around 12 months to come into force.



**Submission: Coimisiún na Meán
Online Safety Code for Video-Sharing
Platform Service**

September 2023

Introduction - Rape Crisis Network Ireland (RCNI)

Rape Crisis Network Ireland (RCNI) is a specialist information and resource centre on rape and all forms of sexual violence. The RCNI role includes the development and coordination of national projects such as using our expertise to influence national policy and social change and supporting and facilitating multi-agency partnerships. We are owned and governed by our member Rape Crisis Centres who provide free advice, counselling, and other support services to survivors of sexual violence in Ireland.

Commentary on Online Safety Code for Video-Sharing Platform Services

The RCNI's focus in this context is prevention strategies combatting rape and sexual violence and providing support for victims who have experienced online violence. In this submission, we will address mostly the harms that we would like to see addressed and unless specifically stated, we would propose a similar overall approach to measures as highlight by the CRA submission.

The gender-based violence perpetrated in an online space presents a particular challenge in developing policies and law to protect victims. Tech-facilitated gender-based violence (TFGBV) is a term defined as:

*'any act that is committed, assisted, aggravated or amplified by the use of information communication technologies or other digital tools which results in or is likely to result in physical, sexual, psychological, social, political or economic harm or other infringements of rights and freedoms.'*¹

TFGBV manifests itself in various ways including: misogyny, discrimination against sex, gender and sexuality, perpetuation of rape myths and victim blaming, coercive control, harassment, stalking, extortion/sextortion, revenge porn, threats, doxing, defamation, impersonation, hacking, hate speech, catfishing, distribution of sexual images and many more equally harmful actions by online users. This form of violence is gendered and has a disproportionate effect

¹ <https://unwomen.org/en/what-we-do/ending-violence-against-women/faqs/tech-facilitated-gender-based-violence>

on women and girls. This form of violence is most commonly perpetrated by men against women in various contexts including intimate partner relationships but also outside of these relationships and particularly against young women who engage in public online spaces such as journalists, activists, politicians and academics.² VSPS provide a mechanism for many of the types of gender-based violence perpetrated online.

For the RCNI the protections needed on VSPS extend beyond only the protection of children or the vulnerable. It extends too beyond the potential victims to include the prevention of the development of potential perpetrators. The effects of online media on the views and perceptions of young people are particularly concerning. Exposure to mass levels of harmful information without sufficient protections and interventions creates cultural baselines that perpetuate negative stereotypes and harmful ideologies relating to sex, gender and sexuality. These cultural biases are shared and disseminated at an extraordinary rate and have a detrimental effect on attitudes towards women and sexual violence. These forms of TFGBV are a precursor to increasing levels of physical sexual violence such as rape and sexual assault as well as an extension of existing violence being perpetrated in intimate partner relationships. VSPS providers are the gatekeepers to an important tool which can also be used as a dangerous weapon. Ensuring that these providers are held to the highest standards and requirements possible is paramount in the protection of victims and the prevention of sexual violence. Having a Code which requires transparency and accountability from VSPS providers is an important step in addressing these cultural influences on sexual violence which are perpetuated and disseminated throughout VSPS.

Submission Questions:

Question 1: What do you think our main priorities and objectives should be in the first binding Online Safety Code for VSPS? What are the main online harms you would like to see it address and why?

² [UNFPA Measuring TF GBV A Discussion Paper FINAL.pdf](#)

The main priorities and objectives of the Code should be transparency and accountability of VSPS providers. The overarching principle should at all times be the protection of users against harm. Not only protection against the harms perpetrated by users but also protection against the harmful effects of content on users directly. From the RCNI perspective, the effects of cultural influences on attitudes towards sex, gender, sexuality and sexual violence as well as the facilitation of gender-based violence on these platforms are the most important harms that we would like to see addressed. The continuum of sexual violence to and from the online space exacerbates the difficulties not only in the protection of victims but also the prevention of these harms.

Question 2: What types of online harms do you think should attract the most stringent risk mitigation measures by VSPS? How could we evaluate the impact of different types of harms e.g. severity, speed at which the harm may be caused? Is there a way of classifying harmful content that you consider it would be useful for us to use?

It goes without saying that any stringent risk measures should be applied to all illegal content but also to all harmful content not yet provided for under legislation. Often the most harmful content can be the most seemingly innocuous. The expression of biases and stereotypes cloaked in freedom of opinion can be hugely damaging and contribute to a culture of hate. Misogyny and attitudes towards sex, gender and sexuality fall within this category. Behind all forms of sexual violence is the belief in the inferiority and inequality of women and by extension the inferiority of what any culture or cohort class as feminine at that point in time. Content that is obvious and easily identifiable is concerning but more damaging is content that is more pervasive and harder to trace. Stringent risk mitigation measures should be applied to both the blatantly harmful content but also to content that is less conspicuous. The evaluation and classification of harm is subjective and therefore difficult to quantify in a way that satisfies all. Cooperation and consultation between stakeholders and regulatory bodies can ensure open communication which allows for constant reassessment of evaluation parameters. Categorisations and general principals can provide a guiding foundation but providing for fast, efficient reporting structures will be most effective in identifying content that is harmful quicker.

Question 3: Do you have reports, academic studies or other relevant independent research that would support your views?

See Annexure 1 attached.

Question 4: What approach do you think we take to the level of detail in the Code? What role could a non-binding guidance play in supplementing the code?

We submit that the Code together with any guidelines or regulations should be binding. VSPS providers are required to submit themselves to regulation in order to do business but are unlikely to take any action which has not been prescribed. If they had more altruistic motivations then these proposed provisions would already be in practice and a Code would not be necessary. There would be little purpose in the Code if the VSPS providers are then simply left to self-regulate. In addition, effective and appropriate sanctions should be included in the Code to ensure compliance by the VSPS providers with its requirements.

Question 5: What do you think would be the most effective structure for the Code? What are the most important factors we should consider when we decide how to structure the Code?

We will not make a specific submission on structure but to say that we consider a combination of detailed provisions together with overarching principles most effective in encompassing a wide range of regulations while guarding against loopholes and technicalities. Factors to consider are firstly, ensuring that the Code's principles cover aspects broadly now but make allowance for new aspects as they arise in the future and secondly that for consumers, regulators and indeed the VSPS providers there are clear and simple thresholds defined upon which action can be taken.

Question 6: How should we design the Code to minimise the potential for conflict and maximise the potential for synergies in how platforms comply with it and the DSA?

We agree with the suggestions made in the Call for Inputs. The Code should minimise conflict and maximise synergy in that it should mirror the DSA at a high-level but provide more detailed instruction and guidance for VSPS providers on how to comply.

Question 7: To what extent, if at all, should the Code require VSPS providers to take measures to address content connected to video content?

Content connected to video content such as comments and attachments can be as harmful and in some cases more harmful when hidden below, embedded or attached to seemingly benign video content. The same stringent risk mitigation measures should be applied to connected video content as that applied to the video content itself.

Questions 8 to 18:

We have no specific submission to make in this regard and refer to the submission made by the CRA.

Question 19: How do you think that cooperation with other regulators and bodies can help us to implement the Code for VSPS?

Consistent and widespread consultation with all stakeholders is essential to developing and maintaining an effective Code of conduct. The use of VSPS is not limited to any one area, sector or country, it is a technology that has multiple uses and multiple effects. Cooperation between regulators and bodies ensures not only knowledge sharing but knowledge production inclusive of diverse viewpoints and experiences.

Question 20: What approach do you think we should take in the Code to address feeds which cause harm because of the aggregate impact of the provide access to? Are there current practices which you consider to be best practice in this regard?

Where users are exposed to large quantities of content which contains harmful messaging, it can result in a belief that either the harmful information is true due to the same information being presented from multiple sources or alternatively that this is the only information available on a particular subject. Specifically when dealing with young people, they are unlikely to conduct independent research to establish the veracity of content they are presented with. We submit that VSPS providers should be required to firstly prevent and control the harmful content but also to put in place measures to ensure that generally feeds contain a mix of content. Furthermore they should have measures in place to flag users whose feeds become dominated by harmful or potentially harmful content to ensure a change to the feed can be introduced to mitigate against its harmful effects. RCNI suggest that algorithms which select content for users based on perceived interests, must obey a 20/80 rule. For example, no matter the commercial or other interest of the platform and its customers, it can only direct a limited percentage of content. The remaining percentage must remain 'free' from algorithmic influence. Child protection limitations such as parental control mechanisms would be exempt to continue to limit content for those purposes. We further suggest that VSPS providers should be required to provide transparency to users as to how they are being profiled. This information should be easily accessible and allow users to correct, alter and control these algorithmic assumptions about them. This user control should be a minimum standard set within the Code.

Question 21: Do you have any views on how requirements for commercial content arranged by a VSPS provider itself should be reflected in the Code?

We do not have any specific submission to make in this regard save that the VSPS provider should be held to the same if not higher standards than those expected of content providers and users.

Question 22: What compliance monitoring and reporting arrangements should we include in the Code?

As stated above, the expectations placed on VSPS providers should be guided by transparency and accountability. VSPS providers should be required to share any and all information required to identify potential risks posed by their systems. Furthermore, they should be required to share potential weaknesses in the protections they have in place or propose to put in place to combat these risks. Compliance statements, while useful for governance, do not always reflect the true picture of whether a system is effective. Internal and external testing of the protections should be required by the VSPS providers in addition to compliance measures. The regulator should have the capacity to audit systems internally and not just be reliant on receipt of volunteered information or information accessed through queries. The code should contain the strict and detailed conditions of such internal access such that all parties can be assured of the protection of commercial and/or sensitive information on the one hand and that access has indeed been facilitated.

Question 23: Should the Code have a transition period or transition periods for specific issues?

While transition periods are understandably necessary, many of the requirements being expected of VSPS providers are extensions or variations on systems that are already or at least should be in place. Any delays in the implementation of the Code and with it the necessary protections results in users being exposed to harms. Any transition periods allowed should only be those that are absolutely necessary and kept to the shortest period possible. That said there will be learning in the roll out of a Code no matter how well crafted. We would suggest that rather than a transition period that the code contains a review mechanism. It is important that such a review mechanism does not become an opportunity for watering down the standards and requirements in the Code. The code review should therefore have strict criteria.

Summary:

The main areas of focus in this submission are the importance of the main objectives and priorities of the Online Safety Code taking into account the following:

- The devastating effects of TFGBV on users and women in particular. Perpetrators use VSPS to harass, coerce and manipulate. This online gendered violence while being a harm in itself, can also be the precursor to physical violence in the form of rape and sexual assault and/or be an extension of existing physical, emotional, psychological and economic violence already being experienced by a victim.
- The damaging effects of content found on VSPS which perpetuates biases and stereotypes that create cultural baselines of hate and prejudice. The primary enabler of the range of harms of sexual violence is the belief in the inferiority and inequality of women and by extension the perception of the 'feminine'. The Code in combating the perpetuation and escalation of sexual violence must take a robust stance against sexism and misogyny.
- The crucial importance of the principles of transparency and accountability being the overarching objective when designing a Code which holds VSPS providers to stringent standards of compliance, monitoring, risk assessment and implementation.

RCNI are at your disposal should you wish to engage with us further on any of these points. While these submissions in their current form are broad, we hope to have an opportunity at a later stage in the drafting process to contribute more specific and detailed information.

Carmichael Centre

North Brunswick Street

Dublin 7 D07 RHA8

September 2023

Web site: www.rcni.ie

Email: legal@rcni.ie

Annexure 1: Research reports

Byerly, C.M. (2020) Incels online reframing sexual violence, *The Communication Review*, 23:4, 290-308, <https://doi.org/10.1080/10714421.2020.1829305>

Crelinstein R. What Can We Do to Combat Online Gender-Based Violence? CIGI

[What Can We Do to Combat Online Gender-Based Violence? - Centre for International Governance Innovation \(cigionline.org\)](https://www.cigionline.org/what-can-we-do-to-combat-online-gender-based-violence/)

Felmlee, D., Inara Rodis,P., & Zhang, A. Sexist Slurs: Reinforcing Feminine Stereotypes Online. *Sex Roles* 83, 16-28 (2020).

Harris, B. , Woodlock, D. Digital Coercive Control: Insights From Two Landmark Domestic Violence Studies, *The British Journal of Criminology*, Volume 59, Issue 3, May 2019, Pages 530–550, <https://doi.org/10.1093/bjc/azy052>

Henry, N., & Powell, A. (2018). Technology-Facilitated Sexual Violence: A Literature Review of Empirical Research. *Trauma, Violence, & Abuse*, 19(2), 195–208. <https://doi-org.ucd.idm.oclc.org/10.1177/1524838016650189>

Tanczer, L. Parkin, S. and López -Neira, I. (2021) ‘I feel like we’re really behind the game’: perspectives of the United Kingdom’s intimate partner violence support sector on the rise of technology-facilitated abuse, *Journal of Gender-Based Violence*, vol 5, no 3, 431–450.

UN Discussion Paper: [UNFPA Measuring TF GBV A Discussion Paper FINAL.pdf](#)

Woodstock, D. (2017). The Abuse of Technology in Domestic Violence and Stalking. *Violence Against Women*, 23(5), 584-602. <https://doi-org.ucd.idm.org/10.1177/1077801216646277>

Coimisiún na Meán (the “Commission”)

In development of the Safety Code for video-sharing platform services

For further information, please contact: Jacinta Brack National Policy & Advocacy Coordinator
The Irish Traveller Movement, 4 - 5 Eustace Street, Dublin 2. 01 679 6577, [REDACTED]

The Irish Traveller Movement welcome the opportunity to submit to Coimisiún na Meán (the “Commission”) on an Online Safety Code.

Founded in 1990, the Irish Traveller Movement is the national advocacy and membership platform which brings together Travellers and representative organisations to develop collective solutions on issues faced by the community to achieve greater equality for Travellers. We represent Traveller interests in national governmental, international and human rights settings. We challenge racism- individual, cultural and structural which Travellers face and promote integration and equality. We are led by our grass roots community membership, deliver expertise in shaping organisations locally and promote community leadership ensuring Traveller’s voices are to the forefront of all discussions.

The Irish Traveller Movement welcomes the new Media and Online Safety regulatory frameworks, the Media and Digital Acts, the establishment of the Media Commission and appointment of a Digital Commissioner for Online Safety. Also, the impending transposing of the Audio-Visual Media Services Directive and the EU Digital Services Act.

Overarching Recommendation for Travellers in the new Online Safety Codes for video-sharing platform services

- Designate Travellers as a protected category in the Codes, to ensure safeguarding and equivalent protection.

The Irish Traveller Movement notes as part of the Commission’s work in developing Codes that it will

- Establish a Youth Advisory Committee
- Conduct research on online harms

We kindly propose that Travellers are involved in the composition and development of both.

In reply we outline matters for Travellers under

- Framing principles of the Code
- Online Harms
- Overall Approach to the Code
- Structure for the Code
- Designing the Code
- Terms and Conditions, Content Moderation and Complaints
- Harmful feeds and recommender systems
- Compliance

Framing principles of the Code

The framing priorities outlined are welcome as such;

- ‘Most importantly, we want the Code to protect children and the public from online harms while upholding and promoting human rights, including the right to Freedom of Expression. And ‘We will take a child-centred

approach to developing the Code where it impacts children’. **The Irish Traveller Movement as a Member of the Children’s Rights Alliance endorse the recommendations made to the Commission in its submission (Sept 2023)**

- In reference to ‘upholding and promoting human rights’ including the ‘right to Freedom of Expression. **There is a robust focus needed to the safeguarding element of human rights within the Codes, where rights are violated under, ‘hate’ and ‘harm and offence’.**

Concerns have been raised that the Online Safety and Media Regulation Act OSMRA did not name Travellers for specific protection which undermined confidence, including for Traveller children. This is problematic, especially as there is still potential for ambiguity/ and a lack of equity generally for Travellers on the basis of hate based commentary. Especially so where it is not understood by services’ moderators, and therefore raised questions as to how hate based harms will be dealt with, unless Travellers are designated for specific protection by name.

We note the Commission intends to ‘meet its obligations’ ‘to ensure that Ireland fully transposes Article 28b of the revised Audio-visual Media Services Directive (the “AVMSD”)’. This does provide confidence where a function of the AVMSD is to ‘Combat racial, religious and other types of hatred by having reinforced rules to combat the incitement to violence or hatred’.

Online Harms

Question 1. Reply: Racism, incitement to hatred and protected categories for ‘at risk groups’

Travellers are one of the most excluded and discriminated groups in Ireland. Online discrimination where ethnic identity is attached to negative reinforcement is very common, and racist commentary widespread. All children and young people deserve to be protected from psychological and emotional harm online, but particular consideration is needed for Traveller children and young people, as video-sharing platform services VSPS most vulnerable users.

Ireland is a signatory to the Framework Convention for the Protection of National Minorities and Travellers are recognised, on grounds of ethnic minority status. It is noted, in view of Article 6 of the Convention ‘Combating hate speech and hate crime’, the Committee in its last report recommend the State party would ‘consider monitoring hate speech in broadcast media as well as online in order to be able to further determine the nature and scope of the phenomenon and to address it, possibly as part of a new national strategy against racism; and establish a mechanism responsible for monitoring social media’

- a) 65% of Travellers in Ireland said they had experienced identity-based discrimination, the second highest finding of 6 European countries researched. And (52%) had the third highest rate of hate-motivated harassment (such as offensive comments on the street or online) (FRA). **(link referenced included below)**
- b) Traveller children are particularly vulnerable as digital natives with increased exposure. Platforms such as TikTok and Facebook facilitate harmful content, by not moderating dedicated sites, including where pages are solely established to either incite or negatively stereotype Travellers, children and young people.
- c) The National Youth Council of Ireland found 79% of participants aged 18 – 24 said racism is a significant issue online.
- d) According to the Yellow Flag Programme, racism as a core social determinant of health inequalities, requires specifically stringent risk mitigation to reduce harm. Defined by Ofcom UK, online racism harms experienced by children and young people can be divided into three psychological impact categories; transient emotional impact, such as confusion, shock or upset, short-term behaviour change, or a more severe emotional impact, such as disengagement with school, long-term harmful behavioural changes, or serious emotional or physical impacts, such as self-harm, complete social exclusion or negative worldview, or view of themselves and their own community. **(link referenced included below)**
- e) Travellers had the highest rate of self-harm acts when compared with other ethnic groups, at (61%) and represented (3%) of 24,473 self-harm and suicide-related ideation presentations at Emergency Departments in Ireland. The peak self-harm age for Traveller men and women, was between 20 and 29 years’. **(link referenced included below)**

Question 2

- **Type:** Racism and Identity hate
- **Evaluation:** Community and ethnic specific harms. Racist content devised solely for spurious purposes.

- **Classifying:** Racist Keyword filtering. Derogatory/ racist slang should be designated out of use, unless for specific defined purposes, where the user has to navigate and authenticate. For example, the word Knacker / Pikey.

Question 3

Hate Crimes. Improvements in hate crime reporting are noted, but largely Travellers have been underreported and are underreporting to racist monitoring, which is influenced by historic discrimination and racial profiling by Gardaí. In 2022 59% of Travellers believed they were stopped by a Garda because they were a Traveller. In 2022 there was also a 29% increase in Garda Síochána recorded hate crimes and hate-related (non-crime) incidents. However, AGS recognised itself these crimes are underreported and the lack of ethnic data including Travellers, is notable. **(links referenced are included)**

This is relevant to proposed Online Safety Codes in these two contexts

1. Underreporting of Traveller's experience as noted above, should not be taken as a basis of extent based on numerical reporting alone, and given there is no specific study of Traveller online racism.
2. Clarity is needed as to how the proposed codes will dovetail with any related criminal enforcement, and how given the experiences outlined to Garda reporting, this might mean Travellers could experience additional barriers.

In reply: Do you have reports, academic studies or other relevant independent research that would support your views? (a-f) as follows:

- a) https://fra.europa.eu/sites/default/files/fra_uploads/fra-2020-roma-and-travellers-survey-country-sheet-ireland_en.pdf
- b) <https://inar.ie/wp-content/uploads/2023/03/Reports-of-Racism-in-Ireland-2022.pdf>
- c) <https://www.ul.ie/news/landmark-study-by-university-of-limerick-researchers-examines-travellers-relationship-with>
- d) <https://www.garda.ie/en/about-us/our-departments/office-of-corporate-communications/press-releases/2023/march/an-garda-siochana-2022-hate-crime-data-and-related-discriminatory-motives.html>
- e) The Office of Communications (Ofcom) in their research paper *Research into Risk Factors That May Lead Children to Harm Online* identified three categories of harm, because of exposure to online harms
- f) <https://www.drugsandalcohol.ie/35870/>

Overall Approach to the Code

Option 3 – A mixed approach (high-level obligations and supplement them with more detail where appropriate)

The Irish Traveller Movement have long advocated for specific safeguarding, given that existing media regulatory frameworks, and online standards and complaints procedures, operate outside the scope of Travellers being considered victims of 'harm' or 'hatred', which is onerous to prove in complaints procedures. The lack of understanding of service moderators to the nature of Traveller racism harm as a basis for complaints.

For example, in June 2023 a music video titled 'Nancy the tinker' produced and uploaded by Dylan Rabbitte-Treacy an arts creator, received 25,000 plays and over 500 likes. It was a singular music output dedicated to deriding Traveller women through a misogynist and derogatory trope. SoundCloud the German music streaming service which enables its users to upload, promote, and share audio, via its user reporting hate speech protocols, commits to 'We will not tolerate content that promotes or encourages hatred, discrimination or violence against others based on things like race, cultural identity or ethnic background, religious beliefs, disability, gender identity, or sexual orientation.'

The Irish Traveller Movement requested the 'video' be removed, SoundCloud's assessment was; 'In this case, the content could be in poor taste. However, there is no clear intention to criticize, or demean any individual or group of individuals on the basis of their belonging to a protected group. This means that we will not be taking further action against the reported user'.

Roma is the term (in European wide human rights equality frameworks) to describe Travellers who are by European and Irish standards, a 'protected' group. **Despite attempts to outline Traveller's ethnic status the German based moderators did not either understand the European standard obligation / and or their own services' observance of categories of protected status.**

Recommendations:

- The code should specify protected characteristics and groups, and ensure a pan European understanding by VSPS.
- Ensure glossaries and guidelines include Travellers and other groups most vulnerable to online harm in the context of racism and hatred.

Effective structure for the Code?

Recommendation The Code should specify metrics on timing and accuracy of moderation decisions and actions in relation to harm, racism and hatred offences.

In reply to: ‘views on how you think we could design the Code to work effectively with other pieces of legislation in the content regulation space, such as the Terrorist Content Online Regulation (TCOR)’.

It is important the Code would also synergise with legislations outside ‘content regulation’ also, and where these will impact on criminal proceeding in the digital space

- 1) The Irish Traveller Movement notes the importance of the forthcoming Criminal Justice (Incitement to Violence or Hatred and Hate Offences) Act, where Travellers are proposed to be specially protected. The Act it is hoped, will be one part of a multi-faceted and comprehensive response, and the Online Safety Code, an important element of that. As a member of the Coalition Against Hate Crime we also **endorse a recommendation for hatred to be defined in that legislation**, via international human rights standards as such; “hatred” means a state of mind characterised as intense and irrational emotions of enmity or detestation against a person or a group of persons in the State or elsewhere on account of their membership or presumed membership of a group defined by reference to protected characteristics, or any one of those characteristics”.
- 2) The European Commission against Racism and Intolerance (ECRI) stressed the importance of drafting (hate crime) provisions in a clear and precise manner within criminal hate speech to ensure legal certainty regarding the scope of conduct that is prohibited, particularly considering the possible interference with the right to freedom of expression. ⁽¹⁾
- 3) Other legislations, where a criminal offence arises in the online space, might also include the Victims’ Rights Directive.

Recommendations

- Travellers should be a named category in the definitions of the Safety Code within a human rights framework.
- The definition of “hate” should be prescribed in the Code as such; bias, prejudice, contempt, hostility and bigotry.

Designing the Code to minimise the potential for conflict and maximise the potential for synergies in how platforms comply with it and the DSA? How should the Code address content connected to video content?

In reply: YES, The Code should address content connected to video content. We agree with the list of ten measures set out in Article 28b.3 of the AVMSD.

Samples of harm noted by Traveller users

Tik Tok

- **Dedicated racist Tik Tok pages** created by fake users and troll accounts of live streams. Filming of family events and dubbed over with racist degrading commentary.
- **Degradation of women:** Sexualisation of Traveller women in videos. Body shaming pages, abusive language.
- **Foul and derogatory language** used against Travellers in live stream videos – “Knacker” “pikey” “dirty smelly Travellers”. Search “Traveller memes” – mostly all negative videos.
- **Lack of consent for racist purposes:** Travellers being used in videos for discriminatory purposes without their consent on Tik Tok.

Facebook

- **Reporting system not reliable.** The word “knacker” can be reported but nothing happen, people are not banned and at times comments don’t get blocked.
- **Pages dedicated to Shame and violence:** shared on the platform or private messaged to others without consent
- **Ethnic profiling.** Traveller activists and in particular Traveller women being targeting online with constant racist abuse and trolling.

WhatsApp

- **Groups established of upwards of 1000 members,** where derogatory and shaming videos are shared, without recourse for people within those videos, from a higher authority, based on usage terms and conditions/regulations.

Recommendations

- Shadow ban or block content creators when engaging in discriminative or racist behaviour.
- Derogatory language should be designated and blocked from the algorithm to prevent searching and sharing
- VSPS should be obligated to train and or hire staff with specification skills and appropriate knowledge of racism, and equality standards.
- The Commission should also replicate that recruitment of trained staff, and convene advisory forums comprising affected groups to inform matters of oversight.
- Video Content. Dubbing, including AI, voice and text overlays should be made identifiable to users on videos.
- User shared videos should be flagged more clearly and designed to allow for harmful themed video content to be picked up by providers which would alert monitoring reviews of that content.

Terms and Conditions, Content Moderation and Complaints

The Commission refer to the current examples of differing standards for pornography, however the area of racism, and incitement to hatred have a higher UN and EU guarantee and **Ireland should ensure our Code places the highest level of protection in this area of harm.**

The Irish Traveller Movement submitted to the Task Force on Safe Participation in Political Life, on the impact of abuse, including online abuse, and harassment, of members of the Traveller community who engage in political life, August 2023. The experiences of Traveller candidates are notable and which we concluded showed a stoical acceptance of the absence of safeguarding and an expectation of being treated less favourably in the digital space, coupled with exhaustive demands of managing and reporting online complaints, complicated by the need to be in digital spaces for discourse on important campaigning topics and to maintain a political profile.

Online harm included:

- a) Dedicated social media bots to look like real accounts and titled under stereotyping names- then linked to the candidate
- b) Constant hatred and personal and family targeting and threatening posts
- c) Racist videos uploaded and dubbed with the candidate’s campaign details
- d) One female candidate was constantly bombarded online with derogatory sexist slurs including, being labelled a ‘whore’, ‘prostitute’, ‘knacker’, ‘drug addict’, and regularly depicted with porn imagery.
- e) In commenting publicly on matters related to identity-based racism and equality issues, levels of hate spiked, and create a ‘pile on’ for trolls.
- f) Over 200 reports were made by one political candidate to social media companies of threatening and or abusive posts and comments, and on online news sites, most resulted in little action.

Onerous burden in reporting Harm

- g) Getting posts taken down where false accounts created in the candidate’s name, required additional evidence to be submitted to social media companies, passport and driver licences

- h) Investigating the trolls and faceless bots, and making Garda complaints, but which could not be upheld due to difficulty tracing people, and worsened where mistrust of the Gardaí, as referred above, the candidates could not be confident of their support.
- i) Counter defeating hate speech and correcting narrative created online about the candidate, necessitating negotiation with traditional media to counter that.

Recommendations

- We agree with the Commission's suggestions and specifically for the Code to include
 - a prohibition on certain types of harmful content such as incitements to violence or hatred, or content which constitutes a criminal offence.
- Ensure VSPS terms and conditions have a universal human rights safeguard in the definition of 'harm'.

Ambiguous Ownership. Even with existing 'harm and offence' standards across VSPS, complaints for Travellers are often not upheld. And where harmful content is shared from one digital platform to another, content ownership is ambiguous.

Recommendation the Terms and Conditions should factor in separate requirements for VSPS from whose platforms harmful content has been shared – i.e.

- platform to platform rules,
- sharing rules and
- content creator rules.

Regarding the proposal (i) 'Establishing and operating transparent, easy-to-use and effective procedures for the handling and resolution of users' complaints to the VSPS provider in relation to the implementation of the measures relating to reporting and flagging, age verification, content rating and parental control systems' **We agree.**

We note the complication as referred earlier – evidenced in ambiguity and lack of expertise of complaints handlers in the areas of racism, and cultural understanding.

Recommendation: Ireland's Code should be robust, so complaints taken are guaranteed to be understood, and have equivalent standing, provide adequate defence from harm and to avoid double harm.

We agree 'users should be periodically reminded of key terms and conditions. These Terms and Conditions should also include glossary terms, for protected characteristics and categories, easily understood by users.

Recommendations

- The Commission undertake an annual audit of VSPS complaints, and include an ethnic identifier be imbedded in procedures, and complaints reporting should be disaggregated based on characterises. (The Committee to the National Action Plan Against Racism, on Ethnic Equality Data outlines the broad equality framing required in Ireland and involving a cross sectoral approach).
- VSPS should be governed by 'sharing' terms and conditions from their platforms, which are clearly identifiable to users, and clearly stating stages the penalty will accrue for users, i.e. point of content origin or from secondary sharing etc.
- Create a centralised VSPS monitoring platform for harmful content. For example; where a Tik Tok generated harmful content video is shared to Facebook and beyond. A systematic red flagging of the content should be shared with other VSPS, and those VSPS have a responsibility to search for and delete from their platforms.

Applying Terms and Conditions

We refer again to the recommendations outlined above and the failure to uphold complaints by companies, as noted by the Commission.

The T&C's for VSPS: There is a need for VSPS to ensure high level expertise for moderators dealing with alleged harm complaints, and for a pan European cultural understanding of groups protected in Article 21 of the European Charter of Fundamental Rights.

Recommendation

- ❑ Adjudication of moderation decisions should be carried out by the Commission, via regular reviews which should be linked (see above) to the complaints reporting (quarterly) and disaggregated for ethnic data.
- ❑ Sanctions for VSPS breaches who consistently under moderate, established via a strict criteria set.
- ❑ Given the need for speed in the digital space, hate / racism complaints should be handled immediately, and a tightening of ‘suspending’ material and content, applied universally where hate based and child-based harm, is defining the complaint raised.

Harmful feeds and recommender systems

We agree the Code should require VSPS to ensure their recommender systems do not result in a feed of content which in aggregate risks causing harm.

For Travellers this is exemplified in many Tik Tok user accounts established to deride, denigrate and cause harm. The platform has a high young Traveller demographic, and since its inception, videos that feature the tag ‘#irishtraveller’ have garnered over 87.3 million views ⁽²⁾.

Regarding the Commission’s suggestion of ‘intercepting a negative feed with positive content’. Concerns arise as to how ‘positive content’ would be objectively decided, and statistically positive Traveller content is also less available. Therefore, safety by design would be a more objective model.

Recommendation

- ❑ Establish an ‘at-risk’ advisory group to work with and inform the Commission’s undertaking of the model, and include content providers

Compliance

It is noted that the Commissioner ‘will have robust compliance and enforcement powers, including the powers to require reporting, initiate investigations and audits, issue compliance and warning notices, and sanction non-compliant online services.

Proof of compliance is essential for Travellers, for confidence in the new Code to bring equivalence and protection not catered for previously. ‘A structured multi-faceted model, where providers give information about their compliance with the Code to the Commission’ is welcome, in addition to ‘an annual compliance statement’.

Recommendation

These compliances should be underpinned with

- quarterly reporting of complaints upheld or not
- ethnic identifier imbedded in reporting
- random and regular investigations of harmful content
- stakeholder / at risk advisory group established and public feedback on risk and harm matters

Regarding ‘approaches if a service’s conduct falls short of that expected by the Code’.

Recommendation

- ❑ Financial penalty, public statement on the service’s platform, advertised publicly, and more robust sanctions for repeat offender services. (However, it is expected ‘approaches’ would be reviewed by the Commission based on reporting outcomes)

- 1) <https://www.coe.int/en/web/european-commission-against-racism-and-intolerance/recommendation-no.15#:~:text=In%20this%20recommendation%2C%20ECRI%20calls,speech%3B%20and%20criminalising%20its%20most>
- 2) TikTok For You page <https://www.tiktok.com/discover/irish-travellers?lang=en>

For further information, please contact: Jacinta Brack National Policy & Advocacy Coordinator and Bernard Joyce Director .The Irish Traveller Movement, 4 - 5 Eustace Street, Dublin 2. 01 679 6577, [REDACTED]

Email: [REDACTED]

**Call for Inputs by Coimisiún na Meán: Developing Ireland's
First Binding Online Safety Code for Video-Sharing Platform
Services**

Submission of the Data Protection Commission

September 2023

Contents

1. Introduction	3
2. Scope of the proposed Online Safety Code	4
3. Age verification	5
3.1 Age verification/assurance in the digital regulation landscape	5
3.2 Age verification, age assurance and age estimation	6
3.3 Age verification under the proposed Online Safety Code	6
3.4 Age verification in a data protection context	7
3.5 Age verification or a “floor of protection”	8
3.6 Taking a risk-based approach to age verification	9
3.7 Age verification methods – Data protection considerations	10
4. Risk assessments and Safety by design	12
4.1 Risk assessments	12
4.2 Risk assessments in a data protection context	13
4.3 Safety by design	14
4.4 Data protection by design and default	14
5. Conclusion	15

1. Introduction

The Data Protection Commission (DPC) is the national independent authority in Ireland with responsibility for upholding the fundamental right of individuals to have their personal data protected, and enforcing the obligations of data controllers and processors in this context. The statutory powers, duties and functions of the DPC are detailed in the Data Protection Act 2018 (the 2018 Act) which gives further effect to the General Data Protection Regulation (EU) No. 2016/679 (the GDPR).

The GDPR, which became applicable as a law on 25 May 2018, recognised for the first time in EU data protection law that there are specific risks posed to children when their personal data is collected and processed and that they therefore merit special protection as data subjects. The GDPR emphasises the need for clear communication with children around how their personal data is processed and points out that children may be less aware of the risks involved in such processing, as well as the consequences of such processing, their rights and the safeguards. As such, the area of protection of children's data and their rights under the GDPR has been a key priority for the DPC since 2018 and is an area in which we are working to substantially raise standards of protection. For this reason, we are pleased to have this opportunity to bring a data protection perspective to this broader discussion on online safety.

While the DPC is primarily concerned with its own area of regulation, namely data protection, the DPC recognises that the regulation of online safety issues and data protection will naturally complement and be mutually supportive of each other. For example, a child's awareness of the risks of sharing their personal data (e.g. posting their phone number or a photo of themselves online) will inevitably support their online (and indeed real-life) safety. As such, the DPC considers that, although online safety issues are outside the remit of data protection law, these objectives are very much two sides of the same coin. It is in this spirit that the DPC wishes to provide some observations to elements of this Call for Inputs from its perspective as a regulator for data protection.

The sections of the Call for Inputs document that the DPC will address in this submission are as follows:

- Section 5.1.3 Age Verification and Age Assurance Features
- Section 5.3.2 Risk assessments
- Section 5.3.3 Safety by design

2. Scope of the proposed Online Safety Code

The DPC notes that this proposed Online Safety Code is being developed in accordance with Coimisiún na Meán’s obligations under the Online Safety and Media Regulation Act 2022 (the “OSMRA”), as well as to ensure that Ireland fully transposes Article 28b of the revised Audiovisual Media Services Directive (the “AVMSD”). This Code is intended to apply solely to video-sharing platform services (VSPS), which are defined in the Call for Inputs document as “a type of online service where users can share videos and engage with a wide range of content and social features”. The DPC notes that this definition includes popular social media services where user-generated videos are available but excludes private messaging. The aim of the Code is to ensure that VSPSs take measures to address online harms more effectively, and the DPC notes that it is also intended that the Code will complement the Digital Services Act (“DSA”) when this comes into full effect in February 2024. Coimisiún na Meán intends for the Online Safety Code and the DSA to complement each other and provide a high level of online safety for everyone.

The DPC wishes to highlight the ongoing work at European Commission level for the development of a Code of Conduct for Age-Appropriate Design under the new Better Internet for Kids Strategy (BIK+)¹, which is also anticipated to tackle the issue of age verification. Therefore, it will be important to ensure consistency with the approach taken at European Commission level.

The DPC welcomes the development of this Online Safety Code for VSPSs, particularly in light of recent statistics published by CyberSafeKids² which state that 84% of 8 to 12 year olds in Ireland have their own social media and/or instant messaging account, and the top four most popular apps are YouTube (76%), WhatsApp (39%), TikTok (37%) and Snapchat (37%). It is evident that children below minimum user age thresholds are actively using many online platforms, and are likely being exposed to harmful or age-inappropriate content³, so it is encouraging to see that three of these top four apps will be captured by the proposed Online Safety Code, in light of their status as video-sharing platform services.

¹ The [new strategy](#) for a better internet for kids (BIK+) was adopted on 11 May 2022

² CyberSafeKids, *Keeping Kids Safer Online. Online. Safety. Matters. Trends and Usage Report Academic Year 2022-2023*. Available at: [CSK Data-Trends-Report-2023-V2-Web-Version.pdf \(cybersafekids.ie\)](#)

³ CyberSafeKids report that 33% of children aged 8-12 years gamed with strangers online, 61% were contacted by a stranger in an online game, and 28% of boys in this age cohort played over-18s game. Further, 26% of 8-12 year olds reported to have seen or experienced something online in the last year that “bothered” them, while 25% have experienced bullying behaviour. In terms of older children, 40% of 12-16 year olds reported that they had experienced bullying online, 26% have seen or experienced something online that “bothered” them, and 40% reported that they post videos of themselves online, 83% of which used TikTok to do so.

3. Age verification

3.1 Age verification/assurance in the digital regulation landscape

The concept of age verification is not new, and has been in use across a variety of sectors for a number of years, primarily in the context of ensuring that under 18s cannot gain access to content or services that are illegal for them to access (e.g. alcohol, online gambling, etc.).

In recent years at an EU level, we have witnessed an increase in the volume of legislation containing references to the concept of age verification, including the AVMSD (which is considered further below) and the DSA⁴. Similarly, the UN Committee on the Rights of the Child also references the concept of age verification in General Comment No. 25 on children’s rights in relation to the digital environment⁵. It is notable that, while the GDPR shines a spotlight on the protection of children’s data, it does not, unlike the AVMSD and the DSA, contain any specific references to age verification, nor does stipulate any specific measures to be implemented (see Section 3.4 for further information).

While technology continues to evolve in this area, there is still no silver bullet or indeed harmonised standard when it comes to age verification/assurance. The suitability of different mechanisms will vary depending on the context and the specific requirements of the legislation or industry in question. However, another important criterion to consider is the purpose or level of certainty to be achieved, for example, does an organisation need to know the **precise age** of a user, do they need to determine if a user is simply **under or over 18**, or more challengingly do they need to know if a user is **over their minimum user age threshold** of 13, for example. As such, the need for confirmation of a concrete age versus a ballpark age is an important distinction as this will have an impact in terms of what the most appropriate solution will be. This is where the distinction between the concepts of age verification, age assurance and age estimation emerges.

⁴ Article 35 (Mitigation of risks) of the Digital Services Act states that VLOPS and VLOSEs shall put reasonable, proportionate and effective mitigation measures in place to mitigate against systemic risk that may occur on these platforms and search engines. Article 35(1)(j) references age verification: “taking targeted measures to protect the rights of the child, including age verification and parental control tools, tools aimed at helping minors signal abuse or obtain support, as appropriate;” Available at: [L_20222777EN.01000101.xml \(europa.eu\)](https://eur-lex.europa.eu/eli/reg/2022/2777/01000101.xml)

⁵ General Comment on children’s rights in relation to the digital environment, OHCHR, March 2021. “Robust age verification systems should be used to prevent children from acquiring access to products and services that are illegal for them to own or use. Such systems should be consistent with data protection and safeguarding requirements.”

3.2 Age verification, age assurance and age estimation

It's important to note the diversification in terminology surrounding the concept of age verification, particularly in the English-speaking world. Several regulators in the digital space, such as the Information Commissioner's Office and the Australian eSafety Commissioner, have defined age verification as a subset of a broader family of methods for ascertaining the age of child users, which collectively fall under the umbrella term of "age assurance". Under this approach, age verification denotes those methods that establish the age of a person with a high degree of certainty (e.g. government-issued ID, electronic identification services, secure third-party services, etc.).⁶ Age assurance, on the other hand, is a term that encompasses both age verification and age estimation solutions. The word "assurance" refers to the varying levels of certainty that different solutions offer in establishing an age or age range.⁷ An age assurance approach allows organisations to select methods that are most suited to the specific risks involved in their processing. Finally, age estimation is defined by the 5Rights Foundation as "a process that establishes a user is *likely* to be of a certain age, fall within an age range, or is over or under a certain age" and includes methods such as comparing the way a user interacts with a device with other users of the same age by testing their capacity or knowledge.⁸ As such, it is important that Coimisiún na Meán be clear in terms of what they mean by "age verification" if and when they suggest recommended measures in this area.

3.3 Age verification under the proposed Online Safety Code

Article 28(b)(3) of the AVMSD sets out a non-exhaustive list of appropriate measures that VSPS providers should implement (as appropriate) in order to protect the general public and children from online harms. One of these measures relates to age verification:

- establishing and operating age verification systems for users of video-sharing platforms with respect to content which may impair the physical, mental or moral development of minors;

The DPC notes from the Call for Inputs document that Coimisiún na Meán is proposing that VSPS providers be "required" to introduce appropriate age-verification mechanisms to protect minors

⁶ See, inter alia, the definitions of age assurance and age verification provided by UK Department of Digital, Culture, Media & Sport in its VoCO (Verification of Children Online) Phase 2 Report: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/934131/November_VoCO_report_V4__pdf.pdf (pp. 12-13), and the equivalent definitions provided by the Australian eSafety Commissioner: <https://www.esafety.gov.au/about-us/consultation-cooperation/age-verification>

⁷ "But How Do They Know It Is A Child? Age Assurance in the Digital World", 5Rights Foundation, March 2021

⁸ Ibid

from online harms in the Code, in light of the fact that Article 28b of the AVMSD requires content that is most harmful to minors to be subject to the strictest access control measures.

Coimisiún na Meán states that some potentially harmful content is inappropriate for all minors, while other content may be suitable for older children but not younger ones. They consider that a VSPS provider may need a system for verifying that a user is an adult or is above a certain age depending on the content that is permitted. As referenced above, this distinction is important as the solutions on the market vary in terms of their ability to determine specific age or estimate an age threshold, so consideration will have to be given as to what kind of age assurance solution is required in a given circumstance.

Coimisiún na Meán is seeking views on whether there are high-risk categories of content that should be subject to the strictest age verification methods (presumably the use of hard identifiers) and if there are lower risk categories that may require a lower order of verification or assurance. While this risk-based approach can work in certain contexts (for example, in the context of data protection which we will discuss below), the content disseminated on video-sharing platforms is, by its very nature, diverse and unpredictable and likely spans the whole spectrum of harmful content, from low to high risk. This would suggest that different age assurance approaches would have to be taken on a video-by-video basis, which seems to the DPC to be an arduous task.

Consideration should also be given as to how VSPSs that do not require sign-up or sign-in to an account to access content (e.g. YouTube) will operate age verification systems. For example, if a child searches for videos without signing into an account, how will the platform ensure that the content delivered is appropriate for that child if they do not know that they are dealing with a child in the first place?

3.4 Age verification in a data protection context

While, as highlighted above, there is no explicit requirement for age verification under the GDPR, Article 8 (and by extension Section 31 of 2018 Act) stipulates that information society services cannot rely on consent as their legal basis for processing the personal data of a child if that child is under the age of 16⁹ in Ireland. If the child is under 16, the organisation must make “reasonable efforts” to ensure that consent has been given by the holder of parental responsibility.

In order to get to a point where an organisation can verify that a parent/guardian has given consent to the processing of their child’s personal data, it may be the case that an online

⁹ Note, the GDPR permits Member States to set a threshold for the age of digital consent between 13 and 16. The age in Ireland has been set at 16.

organisation may first have to ascertain whether the user is a child under the age of 16. As such, while the GDPR does not contain an explicit requirement to verify the age of users in order to identify whether or not they are under the age of digital consent, it may be the case in certain circumstances that this is a practical implication of Article 8.

Articles 24 and 25 of the GDPR can also provide a basis for examining, amongst other things, the issue of age verification/assurance (albeit the concept is not directly referenced). Article 24 focuses on the general obligations of data controllers and requires that controllers must take into account the risks of varying likelihood and severity for the rights and freedoms of natural persons and implement “appropriate technical and organisational measures” to ensure that processing complies with the GDPR. This may come into play where a platform, for example, states that its service is intended for users over a certain age.

Article 25 focuses on the principle of data protection by design and default, and stipulates that controllers must implement appropriate technical and organisational measures which are designed to implement data-protection principles in an effective manner and to **integrate necessary safeguards** into the processing in order to meet the requirements of the GDPR and to protect the rights of data subjects. Controllers must do so taking into account the state of the art, the cost of implementation and the nature, scope, context and purposes of processing as well as the risks of varying likelihood and severity for rights and freedoms of natural persons posed by the processing.

3.5 Age verification or a “floor of protection”

In the context of data protection, it is the DPC’s position that organisations should either take a risk-based approach to age verification/age assurance as mentioned above, or if they do not wish to do this, then they need to apply a **floor of protection** to all users in terms of their processing activities. This means that controllers need to take steps to ensure that all data subjects (irrespective of whether they are under 18 or not) benefit from a high and standardised level of data protection sufficient to protect the rights of any child users.

This is where the purpose of age verification in a data protection context differs from the purpose of age verification from an online safety perspective. Data protection law is not designed or intended to keep users under a certain age off platforms or websites, or to prevent them from accessing a particular type of content; it is designed to ensure that the personal data of these users are protected and that appropriate safeguards are in place. As such, if a controller

can apply this “floor of protection”, there will, in principle, be no need for age verification measures¹⁰ from a data protection perspective.

It is unlikely that this floor of protection approach could be applied in the context of preventing access to certain content, which the Code is seeking to achieve, as to do so would apply a blanket ban on certain content for all users, which is of course neither feasible nor desired.

3.6 Taking a risk-based approach to age verification

As referred to above, the GDPR does not expressly refer to age verification, and equally it is silent on what might be deemed to be appropriate age verification mechanisms.

In practice, from a data protection perspective, many products and services will likely need to rely on a combination of age verification methods in order to ensure the most effective approach. For example, upfront age verification mechanisms such as age gates may only be the first stage in an organisation’s age verification chain, with that mechanism being followed by subsequent steps and interventions which are aimed at building towards a higher degree of confidence about the user’s age. The methods that are most appropriate for organisations will vary considerably from context to context, however, whatever the combination of methods deployed, the result must be **demonstrably robust and effective** and achieve a **level of reliability that is commensurate with the risks** posed by the processing in question. These same principles could also be applied in an online safety context.

In its guidance “[Children Front and Centre: Fundamentals for a Child-Oriented Approach to Data Processing](#)”, the DPC considers a number of criteria that organisations should take into account when taking a risk-based approach to age verification for the purposes of data protection:

- Type of personal data being processed – e.g. health information, images/videos, technical online identifiers, contact details (e.g. full name/age/address/email address/phone number), information about religious beliefs or sexual orientation, information about hobbies or interests, etc.
- The sensitivity of said personal data – e.g. special category personal data, or data which could be considered sensitive for other reasons such as financial information, information on family circumstances or birth status or data which also incorporates the data of a third party such as a family member or friend etc.
- Type of service being offered to the child – e.g. video- or image-hosting platform, educational service, healthcare or social support service, social media app facilitating

¹⁰ This is without prejudice to relevant cases under which Article 8 GDPR may apply.

connections with known parties or with strangers, gaming website, shopping platform, etc.

- The accessibility of the personal data collected to other persons – e.g. whether the nature of the service is to publish or make available the personal data, or elements of it, to the world at large.
- The further processing of personal data including whether data collected is shared with other organisations and the reasons for doing so – e.g. for advertising, marketing or profile-building purposes by either the organisation or any third party with whom the data is shared.

The most stringent age verification methods will always be necessary for online services where the risks arising from data processing or the activities conducted through such services are illegal for children to participate in, for example, where an organisation provides an adult-only service, such as gambling, which by law it cannot provide to under 18s.

While the above-listed criteria are intended to be applied in the context of assessing the suitability of an age verification mechanism in a data protection context, Coimisiún na Meán may find it beneficial to consider similar criteria for a risk-based approach to age verification for online safety purposes.

3.7 Age verification methods – Data protection considerations

There is no one-size-fits-all solution to the issue of age verification/assurance. As mentioned above, appropriate mechanisms are likely to vary from context to context, depending on, for example, factors such as the service being provided and the level of certainty that is required.

In any event, any age verification/assurance measures contemplated for the purposes of online safety should be **proportionate** and grounded on a **risk-based approach**. This means that there should be greater stringency/levels of certainty provided by the particular verification process where the service/content on offer is of higher risk to the user. Any age verification mechanisms developed and utilised, regardless of the purpose for which they are being used, must comply with the obligation of data protection by design and default and must also be subjected to data protection impact assessments in order to assess whether the mechanism in question complies with the principles of data protection under Article 5 of the GDPR:

- Lawfulness – the GDPR requires a **lawful basis** for the processing of any personal data. As such, organisations implementing age assurance measures need to ensure that they have an appropriate lawful basis for doing so under Article 6 of the GDPR. If the use of biometric data is envisaged, organisations will also need to ensure they have an appropriate lawful basis under Article 9 of the GDPR.
- Fairness – the processing of personal data for age assurance purposes must be **fair**. Data should only be processed in ways that people would reasonably expect. The use of age assurance measures should not involve processing data in a way that is misleading or detrimental to the user. No users should be discriminated against as a result of any age assurance measures deployed by an organisation.
- Transparency – organisations should be **transparent** and **up-front** with users about any personal data being processed for the purposes of age assurance, including what personal data is being processed, why it's being processed, who is processing it, how long it will be retained for, and whether any decisions are being made about them as a result of this data being processed. It's also important that organisations explain to users any processes in place for challenging a decision made on the basis of personal data that was processed for age assurance purposes.
- Purpose limitation: Personal data collected for the purposes of verifying age **must not be used by the organisation for any other purpose** (which may entail keeping it separate from other personal data sources which may be used on an ongoing basis e.g. for the ongoing provision of services).
- Data minimisation – The principle of data minimisation requires an organisation to **collect only the minimum information required** to achieve its purpose. When it comes to processing personal data for the purposes of verifying the age of users, there should be no issue with an organisation doing so from a data minimisation perspective, provided the organisation only collects the data necessary in order to be able **to achieve the requisite degree of certainty** about the age of its users i.e. that which is proportionate to the level of risk arising from the processing of personal data.¹¹
- Accuracy: It is very important that any data collected for the purposes of verifying someone's age is **accurate**, as inaccurate data could lead to adults being incorrectly flagged as children, or children being incorrectly flagged as adults. Organisations must monitor and consider carefully any challenges to the accuracy of data.

¹¹ In this regard, it is worth noting Recital 21 of the 2016 Revised Audiovisual Media Services Directive (AVMSD), which recognises that children merit specific protection with regard to the processing of their personal data, and states that the establishment of child protection mechanisms by media service providers *inevitably* leads to the processing of the personal data of minors (emphasis added). Available at: [L_2018303EN.01006901.xml \(europa.eu\)](https://eur-lex.europa.eu/eli/dir/2018/1830/20180626/eng/html)

- Storage limitation: Any **personal data** collected, which provides the basis for the age verification/assurance process to be undertaken, **must be deleted once the appropriate level of confidence as to user's age has been attained**. In this regard, organisations should have clear policies in place as to when and how they consider they have reached such a threshold so that there is a finite point after which the data will no longer be retained.
- Integrity and confidentiality: Organisations must process personal data used for age verification/assurance **securely**. This applies whether an organisation is carrying out the age assurance methods itself or whether they are deploying a third-party solution for age assurance. No matter the circumstance, they should be able to **demonstrate appropriate data security measures and accountability**.
- Accountability: Organisations are responsible for, and must be able to demonstrate, their compliance (through appropriate records and measures) with all of the above-mentioned principles of data protection. In the specific context of age verification/assurance, organisations must be able to demonstrate that their approach to age assurance is proportionate to the risks to users associated with a video-sharing platform service.

4. Risk assessments and Safety by design

In the Call for Inputs document, Coimisiún na Meán includes a section entitled “Additional Measures and Other Matters” which covers additional measures that they may expect VSPS providers to take under the Code, such as following a “safety by design” approach when VSPSs introduce new features, and carrying out risk assessments.

4.1 Risk assessments

Articles 34 and 35 of the DSA will require VSPS that have been designated as Very Large Online Platforms (VLOPs) to prepare systemic risk assessments and to implement risk mitigation measures. Coimisiún na Meán states that there is an alignment between the topics that must be covered in these risk assessments and the risks of harm to be addressed by the Code, and queries whether the Code should require providers of services that are designated both as a VSPS provider and as a VLOP to carry out a similar assessment of the risk of the dissemination of harmful content of the type covered by the Code. Alternatively, they state that they could require a more bespoke assessment of the availability of harmful online content, the risk of it being available and of the risk posed to users, and that this could also include a children's rights impact assessment.

4.2 Risk assessments in a data protection context

Risk assessments are also commonly used in a data protection context and serve as a useful tool for demonstrating compliance with the principle of accountability. Article 35 of the GDPR states that a Data Protection Impact Assessment (“DPIA”) must be conducted by a controller where a type of data processing, in particular using new technologies, is likely to result in a **high risk to the rights and freedoms of individuals**. The GDPR also sets out a number of specific instances in which controllers must conduct a DPIA. A DPIA describes a process designed to identify risks arising out of the processing of personal data and to minimise these risks as far and as early as possible. If required, a DPIA must be completed **prior to** the commencement of the relevant data processing. DPIAs are important tools for negating risk, and for demonstrating compliance with the GDPR.

The GDPR does not explicitly consider the processing of personal data of children to be a processing activity that carries a *high* risk, but the European Data Protection Board’s (EDPB) Guidelines on Data Protection Impact Assessments¹² list “vulnerable data subjects” (to include children) as one of the criteria that could trigger the need for a DPIA.¹³ Additionally, under Article 35(4) of the GDPR, supervisory authorities like the DPC must establish and make public a list of the kind of processing operations which are subject to the requirement for a DPIA. In its published list, the DPC has identified that a DPIA will be mandatory for processing operations involving “profiling vulnerable persons including children to target marketing or online services at such persons”.¹⁴

In its Fundamentals guidance, the DPC considers that the principle of the best interests of the child (upon which the Fundamentals are anchored) under the UN Convention on the Rights of the Child requires that organisations whose services are directed at/intended for children, or likely to be accessed by children, **should** carry out a DPIA in respect of the different types of processing operations which are carried out on the personal data of child users. Such risk assessments should take account of varying ages, capacities and developmental needs of child users as well as considering both actual and potential risks arising from data processing to the

¹² See EDPB Guidelines on Data Protection Impact Assessments (DPIA) and determining whether processing is “likely to result in a high risk” for the purposes of Regulation 2016/679, wp248rev.01

¹³ Data concerning vulnerable data subjects (Recital 75): “the processing of this type of data is a criterion because of the increased power imbalance between the data subjects and the data controller, meaning the individuals may be unable to easily consent to, or oppose, the processing of their data, or exercise their rights. Vulnerable data subjects may include children (they can be considered as not able to knowingly and thoughtfully oppose or consent to the processing of their data)”. (Paragraph 7, page 10)

¹⁴ For more information please see: <https://www.dataprotection.ie/sites/default/files/uploads/2018-11/Data-Protection-Impact-Assessment.pdf>

health, well-being and general best interests of the child, including social, mental, physical and financial harm. The DPC's position is that the best interests of the child principle must be one of the primary risk evaluation tools when carrying out a DPIA concerning the processing of children's personal data.

As part of a child-oriented DPIA, the DPC notes that organisations should consider conducting Child Rights Impact Assessments (CRIA). Prominent academics in the field of children's rights have also highlighted the benefits of using CRIs as a tool "for translating the Convention and its Article 3, on giving priority to the child's best interests, into practice in a concrete, structured manner".¹⁵ A CRIA is a child-focused human rights impact assessment that uses the UNCRC as its framework, and the Digital Futures Commission is exploring the feasibility of digital providers conducting CRIs as a way of embedding children's best interests in a digital world.¹⁶

4.3 Safety by design

The Call for Inputs document states that safety by design involves identifying safety risks in advance of developing a new product or service and considering how to mitigate those risks. Coimisiún na Meán is seeking input as to whether the Code should include a requirement for VSPS providers to publish a "Safety by Design" statement setting out how they consider online safety when developing or enhancing services. They also consider the possibility of including a requirement to prepare a "Safety Impact Assessment" whenever services are being developed or enhanced, with sign-off on the risk assessment and proposed mitigation measures by an executive staff member of the VSPS provider with appropriate experience and responsibilities.

4.4 Data protection by design and default

The sentiments of this principle of "safety by design" are also mirrored in a data protection context under Article 25 of the GDPR, which imposes an obligation of data protection by design and by default on organisations which process personal data. This means that data protection measures should be built into the architecture and functioning of a product or service from the very start of the design process (rather than being considered after the development phase) and

¹⁵ Sylwander, L. (2001). Child Impact Assessments: Swedish Experience of Child Impact Analyses as a tool for implementing the UN Convention on the Rights of the Child (Child Participation). Ministry of Health and Social Affairs, and Ministry of Foreign Affairs, Sweden. See: <https://resourcecentre.savethechildren.net/node/6728/pdf/6728.pdf>

¹⁶ Digital Futures Commission, Pros and cons of child rights impact assessments for digital decision-makers. See: https://digitalfuturescommission.org.uk/blog/pros-and-cons-of-child-rights-impact-assessment-for-digital-decision-makers/#_ftn1

that the strictest privacy settings should automatically apply to a product or service. For example, the user should not have to deactivate (e.g. switch to off) settings which interfere with a person's privacy such as location tracking, health settings which track the movement of a user on a device or settings which automatically broadcast a person's contact details. The DPC considers that these obligations are particularly relevant considerations for organisations whose products or services are directed at/intended for, or are likely to be accessed by children.¹⁷ Recital 78 of the GDPR provides suggested examples of measures which controllers may use as part of their data protection by design and default policy, and the DPC in its Fundamentals guidance has provided a non-exhaustive list of suggested measures that organisations can use to incorporate data protection by design and default to promote the best interests of child users.¹⁸ This includes measures such as turning off geolocation by default for child users, built-in transparency information, implementing parental dashboards (where appropriate), and ensuring limited audience selections by default on platforms where a child can share communications, content or data.

Coimisiún na Meán might find it beneficial to consider drafting a similar list of suggested measures that organisations could take into account in order to incorporate a high level of safety by design into their platforms, products and services.

5. Conclusion

While digital technology and online platforms and services are an intrinsic part of everyday life and provide huge opportunities, the online world also presents new risk scenarios for children and adults alike. As a regulator in the digital space, the DPC welcomes the development of this Online Safety Code for VSPS providers, and hopes that its perspective on the regulation of the processing of personal data online (particularly in relation to age assurance, data protection impact assessments and data protection by design and default) has helped to inform this wider discussion on the prominent issues and challenges relating to the regulation of harmful content and the development of this Online Safety Code. The DPC thanks Coimisiún na Meán for inviting input on these important issues and looks forward to further developments on this Code.

¹⁷ Lievens and van der Hof consider that “[s]ince children are a dedicated category of individuals demanding stricter data protection under the GDPR, the principles of data protection by design and default seem particularly apt to encourage and ensure the protection of their personal data and, at the same time, their rights more generally are guaranteed.” Please see: van der Hof, Simone and Lievens, Eva, *The Importance of Privacy by Design and Data Protection Impact Assessments in Strengthening Protection of Children’s Personal Data under the GDPR (2017)*. *Communications Law* 2018, Vol. 23, No. 1, Available at SSRN: <https://ssrn.com/abstract=3107660>

¹⁸ “*Children Front and Centre: Fundamentals for a Child-Oriented Approach to Data Processing*”, pp 63-66. Available at: [The Fundamentals for a Child-Oriented Approach to Data Processing | Data Protection Commissioner](#)