SUMMARY OF SUBMISSIONS TO CALL FOR INPUTS - “DEVELOPING IRELAND’S FIRST BINDING ONLINE SAFETY CODE FOR VIDEO-SHARING PLATFORM SERVICES”

Prepared for Coimisiún na Meán by CommSol,

October 26, 2023

1 Introduction

This document provides a summary of the responses to the call for inputs (CFI) on the development of a binding online safety Code for video-sharing platform services (VSPS). The summary will constitute part of the accompanying information for the public consultation to be carried out by Coimisiún na Meán ("the Commission") on the Draft Online Safety Code.

A total of 55 responses to the call for inputs were received and the table in section two below categorises responses according to the types of organisations that provided submissions. The categories of types of respondents are listed in alphabetical order in the table and the names of each organisation are listed in alphabetical order within each category.

In some cases, references to the response of any organisation will be identified using the broad categories outlined below. In other cases, responses from one respondent will also include in brackets the names of other organisations that provided a similar response/opinion. This serves to indicate common responses, principles and proposed solutions.

Many of the responses were long and detailed indicating a high level of interest, dedication, and a significant amount of work on the part of those who responded to the CFI. The total volume of responses amounted to almost 800 pages and given the volume and depth of the responses, it was not possible to comprehensively cover all argumentation and evidence provided in every submission. In addition, while some responses did not follow the logic of the questions in the consultation, the very best effort has been made to ensure that all opinions, principles, priorities, and approaches are reflected in this document. This document is intended as an overview of responses for the public to support future consultations on the Code and the aim in producing the final document was that it should be presented in a manner that is clear, concise and understandable for a general audience.

The authors of this report have made all efforts to faithfully and fairly present the main findings and opinions expressed by all respondents in their submissions.

It is foreseen that the majority of the individual responses (dependent upon the wishes of the relevant organisations and individuals) will also be published online. The detailed responses will be further examined by the Commission in the process of developing the Online Code. The availability online also allows for the general public and any other interested parties to delve deeper into the argumentation and evidence provided by respondents.

In addition, many organisations summarised, referenced and provided links to a broad range of research and reports relevant to the issues discussed here. It was not possible in the context of this work to summarise or provide any comprehensive overview of research and evidence provided. Including all the footnotes and references would also have expanded the size of the document. Hence, it is recommended to refer to the individual responses for the detail on research, and to access the reports and links that can be found in the individual submissions of the various organisations.

Finally, the purpose of this summary report is not to evaluate responses, nor provide any conclusions or recommendations (legal or regulatory) or to provide or suggest any technological solutions.

2 Overview and categorisation of respondents to the CFI

Category of respondent	Respondent	Category of respondent	Respondent
Child protection organisations, NGOs, Government agencies and health centres (national and international)	1. 5RIGHTS Foundation 2. Children's Rights Alliance 3. Cybersafe Kids 4. Eurochild 5. Irish Safer Internet Centre 6. National Parents Council 7. Ombudsman for Children 8. St. Clare's and St. Louise's Units (SCU/SLU), Children's Health Ireland	Minority / underrepresented groups and language groups	35. Belong To 36. Conradh na Gaeilge 37. Irish Traveller Movement
Girls and Women's - rights and protection Including also Organisations dealing with violence against women and children	9. Women's Aid 10. National Women's Council of Ireland 11. Rape Crisis Centres 12. Rape Crisis Network Ireland 13. Safe Ireland	News Media	38. News Brands and Local Media
European and international regulatory bodies and government ministries	14. Australian eSafety Commissioner 15. Commissariaat voor de Media - Dutch Regulator 16. Netherlands Ministry 17. Office of Communications (Ofcom), UK	Organisations and NGOs dealing with on online harm	39. Carnegie UK 40. Internet Watch Foundation 41. Trust Alliance Group 42. WeProtect Global Alliance
European co- and self-regulatory bodies	18. FSM – German self-regulator 19. NICAM – Dutch Co-regulator – content classification	Organisations and Government agencies, and research centres dealing with mental health and self-harm issues (suicide, anorexia)	43. Bodywhys 44. Headline 45. HSE National Office for Suicide Prevention (NSOP) 46. National Suicide Research Foundation (NSRF, UCC) 47. Samaritans Ireland
Health promotion organisations providing submissions on advertising	20. Alcohol Action Ireland 21. Breast-Feeding Law Group Ireland (BFLGI) 22. Irish Heart Foundation	Organisations and research centres addressing other harms: bullying, harassment, sharing of intimate images.	48. Dublin City University Anti-Bullying Centre (DCU- ABC) 49. Spunout
Civil rights organisations	23. Irish Council for Civil Liberties	Organisations offering technical solutions	50. Age Verification Providers Association 51. Verify my 52. Yoti
Industry and industry associations	24. Google 25. Meta Platforms Ireland Limited (MPIL) 26. Technology Ireland 27. TikTok Technology Limited	Other individuals	53. SM (anonymous)
Industry organisations providing submissions on advertising	28. Advertising Standards of Ireland (ASAI) 29. Dairy Industry Ireland/ IBEC 30. Food Drink Ireland / IBEC	University departments/ experts/ academics	54. UCD submission 55 Brian O'Neill TU
Irish national agencies, regulators and government departments	31. Competition and Consumer Protection Commission (CCPC) 32. Department of Children, Equality, Disability, Integration and Youth (DCEDIY) 33. Department of Health 34. Data Protection Commission		

3 Online Harms

3.1 Priorities and objectives of the Code and main 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?

Priorities and objectives - Scope of the Code

Several industry responses emphasise that the Code should address the transposition of Article 28b of the AVMSD in Ireland, and therefore focus on the following:

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.

The industry stressed that Article 28b should be the focus as it needs implementation as soon as possible (Industry).

Also, from an industry perspective, the main priorities and objectives of the Commission in relation to the Code should be to ensure clarity, proportionality and the avoidance of duplicate regulatory requirements in the regulation of VSPS.

It is noted that print media and online publications already have policies and practices in place to tackle harmful and illegal online content, and hence it was recommended that the Commission should establish a "baseline" of measures which VSP providers already implement and which form part of industry best practice and existing regulatory requirements. This baseline can then be further built upon where the Commission considers that existing requirements are not sufficient to meet the objectives sought to be achieved by the Code (Meta, Google).

In addition, industry responses call for the Code to not to go beyond the four key areas of the AVMS by extending the measures to ancillary features. 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 (Google).

The Code should reflect and be consistent with the laws governing online services in the EU, in particular the Digital Services Act (DSA), 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 (Google).

It is emphasised that the 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 (Google).

A useful summary outlines four principles that should guide the Commission's Code taking into account relevant differences between services: 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 (Google).

It is also proposed to have a staggered approach: the Commission should focus on AVMSD transposition at this stage and, separately, monitor and assess how the additional harm areas covered by the Online Safety and Media Regulation Act 2022 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 (TikTok).

It is useful to note that in the UK, the Ofcom (national regulatory authority) has been operating a regulatory regime for VSPS for the last two years. This focuses on the AVMS Directive. The Ofcom have not yet introduced a regime related to online harms as the legislation was pending (the Online Safety Bill (OSB) was passed in Parliament in September 2023). Hence, they are currently preparing their approach to harms beyond the AVMS. The Ofcom explains there will be a transition period after the OSB comes into force. "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, as secondary legislation will be needed" (Ofcom). Hence, the feedback from Ofcom for this consultation was largely focused on their experience in the regulation of requirements stemming from the AVMS Directive.

The Dutch Ministry 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). However, the Ministry expressed a strong interest in the broader approach that is taken in the 2009 Broadcasting Act, as amended (Netherlands Ministry).

In relation to audiovisual commercial communications, and the main priorities and objectives for the first Online Safety Code for VSPS and 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 (ASAI).

For other stakeholders, it is stressed that the Code should also incorporate the 42 criminal offences under Irish law listed in Schedule 3 of the 2009 Act as amended, as this covers the most harmful content (WeProtect Global Alliance). The Irish Safer Internet Centre recommends that the harms under Article 28 (b) of the Audio-Visual Media Services Directive (AVMSD) and those outlined in the 2009 Act as amended should 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 (Irish Safer Internet).

According to those organisations that work to protect the interests of the child, the Code should incorporate all the harms listed in the 'Call for Inputs': the four areas set out in Article 28b of the Audiovisual Media Services Directive; the harmful online content relating to 42 criminal offences under Irish law listed in Schedule 3 of the 2009 Act as amended; and content relating to bullying, encouraging behaviour that characterises a feeding or eating disorder, content by which a person promotes or

encourages self-harm or suicide; and content by which a person makes available knowledge of methods of self- harm or suicide (Children's Rights Alliance).

The Ombudsman for Children states that if it is not feasible to prepare a Code that addresses VSPS obligations across EU and national law, there should be a focus on those areas where there is alignment between the Online Safety and Media 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 (Ombudsman for Children).

It is important to note here that under the section regarding priorities, many stakeholders already began to list the types of harms that they believed should be prioritised. In addition, these harms were repeated under the section below, 'Main Online Harms' and also in relation to the question regarding the harms that should require the most stringent risk mitigation measures of platforms. Hence, there is some repetition in these sections.

Main online harms

Priorities of child protection and children's rights organisations

The protection of children from child sexual abuse and exploitation online, the protection of children from bullying and harassment, and from other content harmful to health, well-being and life were top of the agenda for most stakeholders.

A key objective cited by several stakeholders was the need for a Human Rights based approach – particularly focused on the rights of children (5Rights, Alcohol Action Ireland, BFLG, Belong To, Brian O'Neill, Children's Rights Alliance, Department of Children, Equality, Disability, Integration and Youth – DCEDIY-, Irish Heart Foundation). The work of the UN Committee on the rights of the Child and the Council of Europe (COE) 'Guidelines to Respect, Protect and Fulfil the Rights of the Child in the Digital Environment' are regularly cited wherein: 'in all actions regarding the provision, regulation, design, management and use of the digital environment, the best interest of every child is a primary consideration.'

This priority is summarised as follows - the Code should particularly ensure that: the right of the child to protection from abuse and exploitation online is embedded as a key principle; the best interest of every child is a primary consideration in all actions affecting them (Children's Rights Alliance).

Alongside this, it is stated that 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'. It is noted that risks and opportunities associated with children's engagement in the digital environment change depending on their age and stage of development (Irish Safer internet).

The Dutch co-regulator noted that children mainly watch VSPS without parental oversight. Banning illegal content and behaviour from VSPS and providing content information that children of all ages can understand in the blink of an eye is crucial, in order to empower them to choose what they want to watch (NICAM). 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. (Irish Safer internet). A reliable differentiation between children and adults is a key aspect of online safety for children and hence this should be a priority in the Commission's Code (AVPA).

In addition, the OCO encourages the Commission to consider specifying the need for a CRIA (Child Rights Impact Assessment) 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 (Ombudsman). The Department of Children, Equality, Disability, Integration and Youth also emphasised that "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" (DCEDIY).

This is echoed by the Irish Heart Foundation citing the UN Committee on the Rights of the Child - General Comment no.25 on children's rights in relation to the digital environment – "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" (Irish Heart Foundation).

The Online Safety Bill in the UK which passed through Parliament in September 2023 includes obligations for services such as Illegal content risk assessment duties and children's content risk assessment duties (from content linked by the Ofcom in its submission). This in reflected also in the model code developed by Carnegie UK (Carnegie UK).

Priorities of organisations concerned with gender-based violence

Several submissions focused on the priority of dealing with tech-facilitated gender-based violence - 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' (Rape Crisis Network Ireland – RCNI).

According to the RCNI, tech-facilitated gender-based violence 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 (RCNI). There is also a particular concern regarding the non-consensual sharing of intimate images/videos on VSPS, including altered/fake ones, which are becoming more and more common (RCNI, Women's Aid).

Organisations promoting the rights and protection of girls and women and those dealing with violence against women and children outline the following priorities with the emphasis on transparency and accountability:

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: time limits should be put in place for providers to remove illegal or harmful content upon identification. Platforms should be required 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 provider should publish regular reports that include content moderation and enforcement actions.

User Empowerment: the Code should promote collaboration between the providers and educational institutions to promote digital literacy. Providers should be required to promote awareness among users of the avenues of complaint and redress available to them.

Regular Review and Update: the Code should ensure there are regular reviews and updates to adapt to new challenges in the ever-evolving online environments (Rape Crisis Centres, RCNI).

The submission of the National Women's Council (NWC) focused on combatting harms to women and girls in line with the current cross government 'Zero Tolerance strategy on Domestic, Sexual and Gender-Based Violence', and they 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" (NWC, Women's Aid).

Priorities of organisations focused on mental health, mental illness, and suicide.

According to Samaritans Ireland, '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. It can also provide access to content that can be distressing, triggering and instructive and may act to encourage, maintain or exacerbate self-harm and suicidal behaviours. 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' (Samaritans Ireland).

Organisations working in the areas of mental health, mental illness, and suicide have identified the priorities of most concern in their field of work. The Code should ensure that 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. The Code must ensure that VSPS support content moderators and have a minimum standard of care for their content reviewers. If the Commission also intend to develop a 'spot-checking' apparatus, the same minimum standard of care should be applied to Commission staff involved in harmful content review (Headline, Samaritans Ireland).

It is stated that the Code should hold platforms accountable through effective evaluation and monitoring of complaints and reports, made publicly available. It should also build in positive reinforcement mechanisms for VSPS with good compliance (Headline).

The Code should outline compliance monitoring and reporting requirements of content moderators as a keyway 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).

Audiovisual commercial communications and impact on health to be prioritised

With regard to harmful audiovisual commercial communications, the CFI received responses from several organisations focused on the promotion of health (Alcohol Action Ireland, Breast-Feeding Law Group Ireland, Irish Heart Foundation, and the Department of Health).

It was noted that 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. Legal restrictions in the Public Health Alcohol Act do not yet reference the online environment. Hence any new codes which are developed by the Commission must ensure that children are not targeted by alcohol advertisers either in online or traditional broadcast marketing (Alcohol Action Ireland).

Organisations that work in the area of health promotion stated that the main priority and objective of the Online Safety Code should be the protection of children and young people online. Respondents referred to the UN Committee on the Rights of the Child statement that 'the rights of every child must be respected, protected and fulfilled in the digital environment' and the UN Committee recommendation 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. Hence, it should follow that the wellbeing of children must be given primacy over the commercial interests of the alcohol industry (Alcohol Action Ireland).

Some respondents referred to commercial communications and the potential impact on well-being and health of infants - who are not the targets of the advertising. It was stated that harmful Commercial Communications, particularly the marketing of Commercial Milk Formulas, undermine public health, infringe on fundamental rights as enshrined in the Convention on the Rights of the Child and therefore should be addressed as a priority (BFLGI). According to several respondents, the main objectives related to audiovisual commercial communications and impact on health 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 protection in relation to commercial communications; 3. 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; 4. To uphold the best interests of the child as a primary consideration (Irish Heart Foundation, BFLGI).

Addressing deep learning and AI algorithms as a priority

Some academic stakeholders emphasise the danger of algorithms having the ability to decide what content appears on a platform and presenting a profound risk to society. 'The most damaging effects are demonstrated in cases where tragic events such as suicide were linked to content recommended by AI Algorithms. 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' (DCU-ABC). Addressing the issue of algorithmic promotion of hateful and extreme content is one of the key issues to be covered by the Code (Belong To).

The submission of the Irish Council for Civil Liberties has a specific focus on what it terms "digital platforms' algorithmic amplification of hazardous content such as incitement to hate, violence and terrorism, racism and xenophobia" (ICCL). Alongside providing reference to research and recommendations in this area, they cite the Irish Government's National Counter Disinformation Strategy scoping paper of 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" (ICCL).

In terms of priorities, the ICCL provides the following recommendations: “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” (ICCL).

Incitement to hatred, incitement to violence and online discrimination

A key objective for many stakeholders is the provision of measures against incitement to hatred and online discrimination. Research carried out by the DCU Anti-bullying Centre showed that of the adult population in Ireland, just under a half of the respondents 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, in particular, more likely to experience online hate (DCU-ABC).

Spunout emphasised that the Code should take into account hate speech or incitement directed on the basis of gender identity in general and towards trans people in particular. They refer to the significant and well-documented campaigns of harassment and hostility towards trans service users across a number of social media services in recent years (Spunout).

Content that incites violence or hatred against a group of people or members of a group based on any of the grounds referred to in Article 21 of the Charter of Fundamental Rights of the European Union includes language. One of the areas of concern for Conradh na Gaeilge is language, and they stressed that the Commission needs to ensure that there is no discrimination on language matters or any hate speech on that basis, including the Irish language. They state that it is important that 'language' is included, and they recommend that it is taken into account. The Code also presents an opportunity to ensure that there is equality between Irish and English when communicating with the public about online safety for platform services (Conradh na Gaeilge).

The submission from the Irish Traveller Movement emphasises that 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 is widespread. Citing research from the European Union Fundamental Rights Agency, they note that 65% of Travellers in Ireland said they had experienced identity-based discrimination, the second highest finding of 6 European countries researched, while (52%) had the third highest rate of hate-motivated harassment (such as offensive comments on the street or online) (FRA research) (Irish Traveller Movement).

In addition, they note that 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 being video-sharing platform services (VSPS) most vulnerable users. The Irish Traveller Movement as a Member of the Children’s Rights Alliance endorsed the recommendations made to the Commission in its submission and included in this summary. A key issue for the Movement is that Travellers should be designated as a protected category in the Code, to ensure safeguarding and equivalent protection (Irish Traveller Movement).

3.2 Mitigation, evaluation of severity and classification of 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 harm may be caused? Is there a way of classifying harmful content that you consider it would be useful for us to use?

3.2.1 Scope of harms to attract most stringent measures

It is emphasised that 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. This most harmful content relates to the 42 criminal offences under Irish law listed in Schedule 3 of the 2009 Act as amended and would fall under this category (WeProtect Global Alliance).

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 the Code whilst also giving consideration to the provisions of the Digital Services Act (DSA) which could enhance and strengthen the Code (Irish Safer internet Centre).

The greatest and most common types of online harm, especially with respect to video up-loaders, are: content that might impair the physical, mental, or moral development of minors, and; commercial communications including advertising, sponsorship, and product placement, with a specific focus on commercial communications directed towards minors (Commissariaat voor de Media – Dutch regulator).

The industry responses emphasise that while 28b(6) permits Member States to impose 'measures' that are more detailed or stricter than those referred to in paragraph 3, it 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 the Commission, regulated entities and users. They believe that if the Code were to go beyond the requirements of the AVMSD, 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 (Meta).

In order to transpose Article 28b of the AVMSD in Ireland, according to the industry, 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) (Meta, TikTok).

Likewise, Google urge the Commission 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. In addition, the submission states that 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 (Google).

Protection of minors

A key area emphasised by many stakeholders is child sexual abuse and exploitation, which should be a top priority for the Commission (IWF, DCU-ABC, WeProtect Global Alliance, SCU/ SLU of Children's Health Ireland, Cybersafe Kids, Spunout, National Parents Council, RCC). Included in this are the issues

of grooming (SCU/ SLU, Eurochild) and/or inappropriate contact by online predators, and child and human trafficking (IWF, Eurochild, WeProtect Global Alliance, SCU/ SLU of Children's Health Ireland).

According to stakeholders, there have been some disturbing trends identified during the pandemic with some of the key agencies involved in monitoring (including INTERPOL, the National Centre 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 (Cybersafe Kids). It was also noted that there is a growing trend in computer generated child sexual abuse material (RCC). 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.) (Eurochild).

The German self-regulatory body provides evidence from its own research (FSM's "Youth Media Protection Index" study) which shows a significant increase (between 2017 and 2022) in the online harms with which young people aged from 13 and 16 are confronted. These included meeting people online who cannot be trusted (46% of 13/14-year-olds and 60% of 15/16-year-olds) (FSM).

Exposure to pornography and sexual content was also highlighted as a key concern. According to Commonsense Media, the average age at which a child first sees pornography is 12, and 15% of children surveyed had seen it by the age of 10. Research by Childline and by CARI shows that exposure to pornography at a young age can have devastating consequences for the viewer, but also that incidents of peer-on-peer sexual harm increase in populations where children are exposed to pornography (Cybersafe Kids).

In the submission from Ofcom (UK regulator), they stated that their research showed that: "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". For this reason, Ofcom opened an enforcement programme into the adult VSP sector in 2023 (Ofcom). It is also stated, with reference to reports, that many people encounter pornography on Twitter through the accounts of content creators using the platform to drive traffic to their Onlyfans page. As Onlyfans has no advertising, content creators there use other VSPS to promote their sites (NWC).

The NWC also addresses the issue of inappropriate exposure to pornography (most of which is extreme, violent, and degrading to women) at a very young age as being 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, and many believe it to be a driver of sexual violence (NWC, Women's Aid).

Additional harms that were emphasised included video content that promoted "dangerous challenges and stunts.' When the Dutch NICAM were developing their content classification system for the online world, they carried out research with children regarding social media use, experiences with harmful content, and their wants and needs regarding age and content rating. The issue of content featuring 'dangerous challenges and stunts' arose in the research. These challenges are frequently present on social media, and they can pose risks/ dangers to children. According to the research, children do see these quite often and are worried about this content, as are their parents (NICAM). Research from the German self-regulator (FSM) also showed that online harms include being incited to engage in risky behaviour such as dangerous challenges, drug/alcohol use or self-harm (encountered by 35% of 13/14-year-olds and 45% of 15/16-year-olds) (FSM).

Other types of problematic content that arose in the German research included: disturbing or scary content (48% of 13/14-year-olds and 63% of 15/16-year-olds); being the victim of cost traps, rip-offs or scams (27% of 13/14-year-olds and 42% of 15/16-year-olds); being exposed to political or religious extremism (35% of 13/14-year-olds and 49% of 15/16-year-olds) (FSM). The Children's Rights Alliance also quotes European surveys that reveal that children and young people have shown that they are most disturbed by violent content online (Children's Rights Alliance). Concern has also been expressed regarding exposure to extreme violence, horror and torture (including torture of animals) (Cybersafe Kids, Irish Safer Internet Centre, National Parents' Council, NICAM).

Hence, many stakeholders believe that 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 (Safe Ireland).

Hate speech and incitement to violence or hatred

Hate speech was strongly emphasised by several stakeholders (Cybersafe kids, Eurochild, National Parents' Council, RCNI, NICAM, Irish Traveller Movement) including also hateful anti-LGBTQ+ content, and misinformation relating to the LGBTQ+ community and LGBTQ+ (Belong To) and incitement to violence against people or groups of people (Spunout, NICAM). In general, Spunout recommended that the Code should prioritise direct harms such as child sexual exploitation, promotion of suicide or self-harm, incitement to violence against people or groups of people where the harm caused to individuals directly impacts on their life, health or dignity (Spunout).

Gender violence and domestic violence

As noted above that there is concern that the exposure of children to pornography is a contributing factor to harmful attitudes to sex, relationships and gender, and many believe it to be a driver of sexual violence (NWC, Women's Aid). The RCNI emphasise that from their perspective the protections needed on VSPS should extend beyond only the protection of children or the vulnerable. It should extend also 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 (RCNI).

Other important issues emphasised by many stakeholders include the non-consensual sharing of intimate images, and the portrayal of domestic violence (DCU-ABC, Eurochild, IWF, WeProtect Global Alliance, SCU/ SLU of Children's Health Ireland, Women's Aid, RCC, Spunout, Safe Ireland). It was emphasised that the non-consensual sharing of intimate images/videos, should also include altered/fake ones (Women's Aid, Spunout, WeProtect Global Alliance). Other key priorities include image based sexual abuse (IBSA) content, as it is extremely harmful, as confirmed by numerous studies and via the experience of supporting victims/survivors (Women's Aid, Safe Ireland), and Technology-facilitated Gender Based Violence (RCC, RCNI). Also mentioned is 'doxing' described as 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 (RCC, RCNI).

Bullying, harassment and humiliation

Other important concerns include bullying and online harassment and humiliation (SCU/ SLU, Dept. of Health, DCU ABC, National Parents' Council, RCC). 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 - 21% reported dealing with 5+ incidents in that timeframe (Cybersafe Kids). According to recent online safety research on internet use by children in Ireland conducted on behalf of the National Advisory Council for Online Safety, people being nasty to each other (24%), and bullying (22%) stand out as the most mentioned things that upset young people (Irish Safer Internet Centre). They also cite the 'Research Report: Bystander Behaviour Online Among Young People in Ireland' (DCU ABC), which shows, among others, that cyberbullying is a significant issue encountered online by children in Ireland. The DCU anti-bullying Centre also states that their research shows that children have various understandings of what cyberbullying and harassment are. Hence, the Code should take into account the relevant definitions of the phenomena and how social media platforms operationalise them for moderation (DCU ABC).

Research from the German self-regulator also showed that being bullied by others online (encountered by 51% of 13/14-year-olds and 53% of 15/16-year- olds); and being harassed online (encountered by 51% of 13/14-year-olds and 56% of 15/16-year- olds) were significant harms (FSM).

Suicide, self-harm and feeding disorders

Content that promotes self-harm or suicide, or promotes eating disorders was highlighted by several stakeholders as an area that should be prioritised (Cybersafe Kids, Eurochild, HSE NOSP, Dept. of Health, Bodywhys, Headline, SCU/ SLU, DCU ABC, Spunout, NICAM). This should include content promoting and encouraging behaviour that characterises an eating disorder (DCU ABC, Dept. of Health).

As noted above, certain stakeholders emphasise that VSPS need to be 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 (Headline, Samaritans Ireland).

The National Suicide Research Foundation at UCC stated that the most stringent measures should be taken to deal with a range of content such as: content on online information sources such as websites and platforms used to inform people on methods for suicide; similar content on search engines; similar content on social networking sites; also online imagery and videos; online forums or message boards; pro-suicide and self-harm websites; online suicide 'games'; the 'Darknet'; Livestream suicide/Cybersuicide; and online Suicide 'pacts' (National Suicide Research Foundation - UCC).

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 development of the online safety Code. 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. They also state that the Code 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 (HSE NOSP).

According to Headline, 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).

Regarding self-harm and eating disorders, the HSE National office for suicide prevention (NOSP) highlights that eating disorders have the highest mortality and morbidity of all of the mental health conditions, and it is estimated that they will affect between 1–4% of the population at some point in their lives (HSE-NOSP).

Stakeholders also noted the problem of the context in which the content is presented. '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' (Dutch Ministry). Online harms can be emphasised or increased by frequency of occurrence of topics in a child's media menu. This is controlled by algorithms, especially where VSPS 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 normalisation. Eating disorders, self-harm, etc. are subjects where big risks can currently be seen. This contextual algorithmic factor should be addressed in the Code. Technical execution would vary from VSPS to VSPS but in general terms stakeholders would recommend adopting an 'algorithm after consent' approach. This would imply that the platform is fit for all unless indicated that you are an adult (account) or have an adult's consent (NICAM, Dutch co-regulatory body). Bodywhys stresses the growing incidence of eating disorders - citing media reports (2022), which identified an almost five-fold increase in cases of eating disorders at the Children's Hospital in Tallaght over the past eight years. In July 2023, the Health Research Board (HRB) reported that the number of child and adolescent admissions for eating disorders more than doubled in the last 5 years, from 33 in 2018 to 80 in 2022 (Bodywhys).

Amplification of harm

The Irish Council for Civil Liberties submission (as outlined above) focused on digital platforms' algorithmic amplification of hazardous content such as incitement to hate, violence and terrorism, racism and xenophobia. They state that recommender systems are understood to be dangerous and require prioritisation. 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. Further, they state that the content covered by section 139K(2)(c) of the Online Safety and Media Regulation Act 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 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. As noted above, they emphasise that acting against algorithmic amplification rather than attempting to identify and un-publish harmful content is likely to be more effective and avoids intrusion upon the right to freedom of expression (Irish Council for Civil Liberties).

The SCU/ SLU also emphasised the problem of algorithmic targeting of certain content such as content that is related to self-harm/suicide/eating disorders, to child sexual abuse material, and inappropriate

adult content (SCU/ SLU). Belong To also note research by organisations such as Hate Aide, that found that 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 (Belong To).

Advertising and harms

Certain types of audiovisual commercial communications and advertising were considered to require stringent risk mitigation measures (Breast-Feeding Law Group Ireland - BFLG, Irish Heart Foundation, Children's Rights Alliance, Alcohol Action Ireland).

According to the responses of a range of stakeholders, Harmful Commercial Communications, particularly the 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 in the Code (Breast-Feeding Law Group Ireland - BFLG, Irish Heart Foundation).

In addition, the advertising of alcohol is emphasised as being of risk to the health and future well-being of children and young people. Alcohol Action Ireland cited a recent World Health Organization report, which found 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" (Alcohol Action Ireland).

The fundamental concern of these organisations was the impact of advertising, marketing and commercial communications on children's health and their long-term health and well-being including life expectancy. In addition, it was emphasised that the Code should protect all children regardless of age or their ability to use digital media (Action Alcohol Ireland, Breast-Feeding Law Group Ireland, Irish Heart Foundation). This is particularly pertinent in relation to breastmilk substitutes or commercial milk formula, which are marketed to mothers and can have immediate or life-long impact on the health of children (Breast-Feeding Law Group Ireland).

The Irish Heart Foundation (IHF) notes that a number of the risk factors for cardiovascular disease 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 requires concrete action to be taken to protect children from exploitation and harms. 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 (Irish Heart Foundation).

The Irish Heart Foundation (IHF) presents a range of evidence related to advertising and marketing of HFSS Foods, Commercial Milk Formula and E-Cigarettes. They echo 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 (Irish Heart Foundation).

In a 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. Experts claim that 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" (Irish Heart Foundation).

According to BFLG, 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 commercial milk formula 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 commercial milk formula brands' digital platforms to provide parents with parenting information (Breast-Feeding Law Group Ireland – BFLG).

The Department of Health emphasises that the OSMR Act (in section 139k(5)) states that codes and rules may prohibit or restrict the inclusion in programmes or user-generated content of commercial communications 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 those foods or beverages which contain fat, trans-fatty acids, salts or sugars (Dept. of Health).

The Department of Health also referred to the 'Healthy Ireland- Obesity policy', which is the policy instrument for obesity in Ireland - "A Healthy Weight for Ireland'. The Obesity Policy and Action Plan (OPAP) was launched in September 2016. The establishment of the Commission with a remit to develop 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 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. The standards and practices that can be addressed through regulatory codes and rules developed by the Commission should include the advertisement of certain foods and beverages. 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 (Dept. of Health).

The Dairy Industry Ireland (DII) stated that they strongly supported 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. They note that regulation of such communications is already set down at Irish and EU level (Commission Delegated Regulation (EU) 2016/127, Regulation (EU) No. 1169/2011 on Food Information to Consumers, Regulation (EC) no 1924/2006 of the European Parliament and of the Council on Nutrition and Health Claims made on Foods, 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).

They also state that the 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, critically reflects the already stringent EU regulatory frameworks that govern communication in relation to formula milks. The industry is willing to engage with regulatory authorities and ensure strict compliance with the law relating to product communication. The DII fully agree that breastfeeding is the best source of nutrition for babies and should be promoted and protected, with all necessary supports in place. 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 (Dairy Industry Ireland - DII).

Food Drink Ireland (FDI) strongly voice their 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.

The FDI also make reference to the Best-ReMaP, which 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. The aim is to reduce children's exposure to the marketing of foods high in fat, salt and sugar (HFSS), including online. However, 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 the Commission with a full suite of powers. The FDI state that 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 the Commission (Food Drink Ireland - FDI).

Evaluating the impact of different types of harms e.g., severity, speed at which harm may be caused

Several responses provided suggestions regarding the evaluation of impact. For example, it was suggested that in order to evaluate the impact several 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 normalise unrealistic body types or sexualised images of young women and girls.

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.

3. Direct and indirect - The way in which a child is interacting with the content can influence the impact. 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 directly or indirectly the impacts of harm are felt (5Rights)

In order to evaluate the impact of these harms SCU/ SLU recommend that the Commission creates links with organisations who work with vulnerable populations to seek quantitative/qualitative information. This also represents an opportunity to create research partnerships to broaden knowledge, expertise and ultimately build effective responses to harms posed in this space (SCU/ SLU).

The Dutch co-regulatory organisation (NICAM) indicated that they have been working on a scientific approach towards standardised 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 should not be exposed to certain content. The objective, standardised 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 (NICAM).

Evaluating the impact of different types of online harms involves assessing various factors, and the HSE NOSP specified a range of 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, for example, 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 be bereaved by suicide. The design of codes should consider 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 (HSE NSOP).

There should also be scope to review the real-world impact of such harms, i.e., by 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 take into consideration 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 (HSE NOSP).

According to the National Suicide Research Foundation, the 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) (NSRF-UCC).

Is there a way of classifying harmful content that you consider it would be useful for us to use?

The Children's Rights Alliance indicated that 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' (Children's Rights Alliance). They also suggested a number of frameworks that could be considered in this regard. The first is the 4Cs classification.

The CO:RE 4Cs classification is a key tool to identify risk and classification of harm is the 4Cs framework. The 4Cs classification recognises that online risks arise when a child:

- Engages with and/or is exposed to potentially harmful **c**ontent,
- Experiences and/or is targeted by potentially harmful **c**ontact,
- Witnesses, participates in and/or is a victim of potentially harmful **c**onduct,
- Is party to and/or exploited by a potentially harmful **c**ontract.

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 (Children's Rights Alliance, referenced also by Belong To, Bodywhys). This system 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 risks to the values that shape childhood and society are increasingly prominent (Irish Safer Internet Centre). 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 (Irish Safer Internet Centre, DCU-ABC). Mention was also made of the Australian Classification Scheme, the UK Classification Scheme (Children's Rights Alliance), and the UK Online Safety Data Initiative's Taxonomy project (DCU-ABC).

The Better Internet for Kids report was cited as providing a helpful framework for classifying online risks and subsequent harms. Based on the EU Kids Online research framework, 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, and there is further research that has attempted to classify severity of harmful online content (DCU ABC).

The Australian Classification Scheme was cited by Brian O'Neill as an illustration of 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). 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 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 (Brian O'Neill).

The Children's Rights Alliance recommended that the Code should:

- Provide for mechanisms for child sexual abuse material to be removed swiftly;
- Include measures to address children's access to pornography and the advertising of prostitution;
- Provide for measures so that illegal material such as child sexual abuse materials, intimate images and material that incites hatred can be robustly and swiftly removed;
- Address violent content online (Children's Rights Alliance).

Eurochild noted that instead of using narrow categories such as "self-harm" or "sexual harm", that the 4Cs classification allows for the discovery of the different manifestations 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 harmful content. These two require very different responses. This classification can enrich the risk assessment and mitigation exercise (Eurochild).

It was emphasised 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 there is a need to refer to different groups of children or address stakeholders;
- 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, one should 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);
- 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 (Irish Safer Internet Centre).

The SCU/ SLU recommended consideration of the following classifications of harmful content (not listed in order of priority): mental health; sexually harmful/abusive material; adult sexual material; bullying and harassment; disinformation; extreme violence and aggression (SCU/ SLU).

The Irish Traveller Movement emphasised the need for racist keyword filtering. Derogatory/ racist slang, for example, the use of words such as Knacker / Pikey should be designated out of use, unless for specific defined purposes, where the user has to navigate and authenticate (Irish Traveller Movement).

The Department of Health reported that Ireland is currently working with European partners on a Joint Action called Best-ReMap 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 of unhealthy or harmful foods and beverages to children. This work is due to conclude shortly. Further engagement with the Commission is foreseen by the Department and recommendations will be determined by the outcomes from the EU Joint Action on Best-ReMap (Department of Health).

On classification of harmful content, the NSRF recommended categorising the following types of online content in line with international research (McTernan & Ryan, 2023; Susi et al 2023, Marchant et al. 2017). Such content is retrieved via various sources: (1) Online information sources (websites and platforms) and (2) Search engines; 3) Social networking sites; 4) Online imagery and video; (5) Online forums or message boards; (6) Pro-suicide and self-harm websites; (7) Online suicide 'games'; 8) The 'Darknet'; (9) Livestream suicide/Cybersuicide; (10) Online Suicide 'pacts'. Among others these: facilitate access to potentially harmful information; facilitate contagion; normalise self-harm and suicide; increase risks following celebrity suicide; facilitate cyberbullying; share suicide notes (NSRF-UCC).

It is proposed that content should be classified in the following may: to prioritise illegal content (offence-specific categories of online content), as defined in national and EU law, which should attract the most stringent risk mitigation measures; to give effect to measures addressing other harmful content that meet the high bar of the risk test in the Online Safety and Media Regulation Act 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; and regarding additional content such as that which illustrates bullying or humiliating, or 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 - should all be subject to risk mitigation measures in accordance with the likelihood of access and potential impact on vulnerable subjects (Brian O' Neill).

3.3 Reports, academic studies or any other relevant independent research

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 considerable number of reports, academic studies and other relevant research, tools, guidelines and recommendations were provided by respondents, and these are detailed in the individual responses with links to various documents. Additional reports were provided by Industry actors such as Community Guidelines Enforcement Reports (Google), Meta reports on various codes and best practice, and reference to the research of other regulators.

4 Overall Approach to the Online Safety Code

This section provides an overview of the responses to questions 4, 5, 6 and 7.

4.1 Level of detail of the Code and the role of non-binding guidance

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 the three options on how prescriptive the Code should be, the majority of stakeholders call for a middle way e.g., a mixed approach. A middle way would impose high-level obligations and supplement these with more detail where appropriate. For example, the Code could include a high-level obligation that requires VSPS providers to have and to apply terms and conditions that prohibit users from uploading hate speech videos. The Code could then further specify that hate speech videos (or other videos) must be taken down within a specified time after being flagged. The Code might be supplemented with more detailed guidance to assist platforms with compliance. The VSPS providers would be required 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.

A mixed approach

Several stakeholders share the opinion that setting high-level principles will allow for innovation in terms of good practices and will ensure the Code remains technology neutral and future proof, while the specific commitments will ensure the Code is actionable and measurable (Irish Safer Internet, The Internet Watch Foundation, DCU-ABC, FSM, 5Rights foundation, Eurochild, AVPA, Brian O'Neill).

According to the DCU response, a very detailed, prescriptive code may fail to take differences in the 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, alternatively, allow companies to evade adequate scrutiny (DCU-ABC).

Safe Ireland's view is that a brief statement of a high-level obligation, should be 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. Regarding non-binding online safety guidance materials, they emphasise the importance of context: a single piece of online material which could be relatively harmless in one context might be very damaging in the context of a pattern of domestic abuse (Safe Ireland).

The Department of Health shares the opinion that highly detailed or overly prescriptive regulatory frameworks can risk a 'letter' rather than 'spirit' approach from those being regulated and encourage a narrower focus, as opposed to 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 (Dept. of Health).

The Internet Watch Foundation (IWF) favours a mixed regulatory approach as this gives regulators the ability to be flexible in their response in order that regulation continues to keep pace with changes in technology. Many of these platforms are unique in the way they are designed, and the platforms are established in different ways, despite the fact they may appear to have very similar characteristics. The IWF 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 (IWF).

The HSE NOSP (National Office for Suicide Prevention) view is that a mixed approach with non-binding guidance is the best way to approach the level of detail required in the Code. 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. 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. Similarly, the Samaritans would like to see all platforms adopt co-designed guidance, such as their 'Online Safety Industry Guidelines for managing self-harm and suicide content' (HSE NOSP, Samaritans Ireland).

The Dutch regulator (Commissariaat voor de Media) also supports a mixed approach with 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. 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, but not in the way it is implemented (CvdM).

A very detailed, prescriptive Code

The second largest group of stakeholders proposes a Code with detailed and descriptive measures for VSPS providers in addressing online harms.

Cybersafe Kids believes that a flexible code is akin to self-regulation, and this does not work because these measures will be contrary to the natural inclination towards commercial benefit for the companies. They claim this is the reason why legislation had to be introduced. This Code will be the basis on which penalties will be imposed so it needs to be very specific, clear and prescriptive. The companies need to be clear when they are breaching the Code, otherwise they will not know how to adhere to it. Additionally, they propose to take different approaches when 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 Online Safety Commissioner via the Individual Complaints Mechanism (Cybersafe Kids).

According to the SCU/SLU, a very detailed Code is necessary due to a history of a lack of interest on the part of online technology companies in addressing concerns regarding safety and an increase in the impact of online harms. 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 (SCU/ SLU).

From the experience of Women's Aid, women contacting platforms to have material taken down can find this to be a frustrating and traumatising experience, 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. Non-binding guidance for platforms are welcome to help ensure consistency and clarity but there needs to be enforcement of the Code. Binding rules are therefore more important

(Women's Aid). A similar opinion is shared by the Rape Crisis Centres and the Rape Crisis Network Ireland adding that VSPS are unlikely to take any action which has not been prescribed. They emphasise that if the companies 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 (RCC, RCNI).

The Dutch co-regulator 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 (NICAM).

The Breast-Feeding Law Group of Ireland (BFLGI) supports a detailed prescriptive Code with evidence that the Commercial Milk Formula industry has, in their view, 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 Code represents 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. The sophisticated digital marketing deployed by commercial milk formula brands needs careful, considered and exacting monitoring and regulation and cannot be left to industry (BFLGI).

The Irish Council for Civil Liberties (ICCL) believes voluntary and discretionary measures by platforms will not be sufficient. Therefore, they are supporting a very detailed prescriptive Code. The current voluntary measures of Digital platforms against the risks 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". It concluded that 'Article 35 [DSA] standards of effective risk mitigation were not met in the case of Kremlin disinformation campaigns' (ICCL).

Headline claims that their experience demonstrates that the Codes should include as much detail as possible. For instance, 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 Code must be robust, it must also provide enough detail to the audience, the uploader, and the VSPS to avoid misinterpretation of the intended protections (Headline).

A very high-level code

The industry proposes a high-level regulatory practice as it leaves 'maximum flexibility as to how goals can be achieved'. The Code must set clear goals but enable VSPS to deliver upon those goals based on the specific characteristics and functionalities of their service. To achieve this, they believe that the Commission should implement a Code that is proportionate, risk based and focused on the principle of co-regulation. A detailed and prescriptive Code – with rules on terms of service and the design of products would: 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'; 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, limit 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 (Google).

Google notes that a non-prescriptive approach to the implementation of the AVMS Directive reflects the approach adopted in many other Member States, and any deviation from this approach should be carefully considered. In a recently published report by the Ofcom on the terms and conditions of VSPS registered in UK they 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' (Google).

A similar opinion was expressed by TikTok stating that they operate under the UK's AVMSD implementation have found that it has worked well in practice. They also 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 that VSPS are already subject to under existing regulations. Additionally, the more prescriptive the measures the more likely they are to be unworkable for certain VSPS and therefore ineffective, as one size will not fit all. 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 (TikTok).

Therefore, the industry responses stress that the Code should be based on principles 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 (Google).

Meta shares a similar opinion that the new Code should establish a "baseline" of compliance measures for VSPS and a method of monitoring the effectiveness of that baseline. They also believe the Code should align with established regulatory regimes (such as the DSA and the Terrorist Content Online Regulation (the TCO Regulation)). Regarding non-binding guidance, the AVMSD explicitly encourages the use of self- and co-regulation to meet its objectives. They would encourage the Commission 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 the Commission 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 may also draw up additional voluntary codes of conduct to tackle different types of systemic risks and illegal content (Meta).

In addition, it is emphasised that there is a clear risk that a prescriptive Code with detailed obligations on VSPSs may run the risk of discouraging or disincentivising VSPS 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 (TikTok).

4.2 Overall structure of 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?

- To have separate sections in the Code for each main category of content it addresses.
- The structure of the Code may be thematically based on how the different parts of VSPS are impacted by the appropriate measures set out in Article 28b.3 of the AVMSD. For example, the Code could be split into the following sections: Content Policies / T&Cs; Risk Assessments; Content Moderation and Complaints; Online Safety Features; Service Design Measures; Compliance Measures.
- The structure the Code may follow the Article 28b (3) measures of the AVMSD sequentially from (a) to (j).
- In case of a very high-level code, it may only have one or two sections.

The Code structured in separate sections.

Several stakeholders support the proposal that the Code be structured in separate sections. The SCU/SLU proposes the use of simplified language to ensure that the Code cannot be easily misinterpreted. This might require the provision of supplementary documentation if necessary (SCU/ SLU). A similar view is shared by Women's Aid and proposes a separate section for each main category of content making the Code specific and clear regarding how each harm is addressed. A key separate section should be 'image based sexual abuse'. Each major section should have similar subsections addressing the relevant measures (Content Policies / T&Cs; Risk Assessments; Content Moderation and Complaints; Online Safety Features; Service Design Measures; Compliance Measures) (Women's Aid).

According to the Department of Health a factor to consider in structuring the Code would be the need to list categories of harmful content, but also be specific in defining what types of content can fall within a particular category, for instance: Suicide and self-harm content includes 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, and content about celebrity suicides, which can increase risk. The code should 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 to be taken by the relevant platforms (Dept. of Health).

The DCU-ABC recommends following the Australian Online Safety Commissioner's Guidelines for the industry associations. They broadly outline objectives with specific outcomes under each objective, and compliance criteria per outcome. DCU-ABC suggest the Commission might adopt similar best practices when formulating something similar for each category or several categories of harmful online content, and they reiterate their position on the importance of distinguishing between online risk and harm (DCU ABC). A thematic approach is considered the most appropriate and it is recommended to consult the models developed by the Australian eSafety Commissioner (Brian O' Neill).

The Irish Safer Internet Centre believes structuring the Code thematically will allow VSPS providers 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 ensure it is 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 (Irish Safer Internet Centre).

The Code structure is thematically based on how the different parts of VSPS are impacted by the appropriate measures set out in Article 28b (3) of the AVMSD.

One response from the industry proposes that the Code should 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. They also believe it may be useful for the Code to set out any guiding and overarching principles at the beginning. In their view, one of the most important factors to be considered when designing the structure of the Code is adaptability. For the Code to be effective, it is crucial that it 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 be future proof (TikTok).

The Code structure is based on Article 28b (3) measures of the AVMSD sequentially from (a) to (j)

Some stakeholders are in favour of the Code structured in terms and items of the AVMSD, for example the 10 measures provided by Article 28b (3) and from the DSA where possible. Eurochild stresses that 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 (Eurochild). A similar approach was recommended by the German self-regulator (FSM) stating that 'the Code could be structured along the Article 28b (3) measures of the AVMSD' (FSM).

A very high-level Code with only two sections

Industry stakeholders propose a very high-level Code and foresee only one or two sections. 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, allowing for greater adaptability and proportionality, and accordingly more effective, future-proofed regulation of VSPS (Google).

Meta believes the Code will be most effective if it is principles based. Accordingly, the Code should generally identify the types of harmful content that it seeks to address and should set out the obligations applicable to VSPS, in general terms. They emphasise that the Code should not prescribe obligations considering categories of content, as, often, the same types of measures are appropriate to deploy across different harm types. Regarding the structure. 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 (Meta).

Mixed approach to the Code structure

The Rape Crisis Network Ireland considers a combination of detailed provisions together with overarching principles most effective in encompassing a wide range of regulations while guarding against loopholes and technicalities. In their opinion, the factors to consider are firstly, to ensure 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 (RCNI).

Similarly, Safe Ireland's view is that it would be most useful to combine the first two approaches listed in the Call for Inputs document, 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 28b (3) of the Directive, but perhaps grouping measures related to prevention together instead of in the order in the document. Thus, in this order - Content Policies/Terms and Conditions, Online Safety Features, Service Design Measures, Risk Assessments, Content Moderation and Complaints and Compliance Measures. In their 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). Safe Ireland also suggests 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).

4.3 The Digital Services Act- preventing conflict and maximising synergies

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 responses related to this question covered several areas, which partly relate to the issues addressed above: to what extent a more detailed outline of what is understood by harms would go beyond the principles-based approach of the DSA, or on the other hand the extent to which the Code must articulate the harms; the concern of the industry that conflicting regulatory regimes may be established; and discussions on the potential for positive synergies between the Code and the DSA particularly regarding the implementation of both.

The majority of stakeholders are of the opinion that the Code should impose more detailed requirements on VSPS providers than that which is outlined in the DSA, including with regards to the privacy, safety and security of minors that are specific for VSPS. Examples of key issues already outlined above include the following: age assurance/verification, content rating, profiling of minors by recommender systems or parental controls, the use and design of recommender systems, moderating content including notice and takedown, and addressing specific gendered violence and harm that can be affected and perpetuated against women and girls (5Rights).

It is recognised that additional requirements should not conflict 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 (5Rights).

The Online Safety Code needs to be compatible with Ireland's International and Domestic Legislation. Of particular importance are the UN Conventions including the Convention on the Rights of the Child, the Irish Constitution and the European Convention on Human Rights. It is important that the Commission ensure that when designing the Code that particular attention is paid to ensuring it is compliant with international and national human rights law (Children's Rights Alliance).

The Irish Digital Services Bill 2023 and the EU Regulations under the Digital Services Act 2022 should be seen as a collaborative resource to further enhance and strengthen the Code, rather than one that has the potential for conflict (SCU/ SLU). The SCU/SLU further recommend that the Code is aligned with these frameworks, to ensure maximum synergy, thus promoting the safety needs of individuals accessing VSPS, while also holding these platforms to account.

The Irish Safer Internet centre points out that the Code will have to consider how providers subject to the DSA will have to comply with definitions of online harms in other national legislation, such as provided for in the 2009 Act as amended. Providing there is no synergy, 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. While 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. They also see opportunities for synergy between the proposed super complaints mechanism in the 2009 Act as amended and the proposed trusted flagger scheme in the DSA, as they 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 systematically. The objectives of both could be aligned. Other synergies could also be found in how the Code addresses online harms.

A further synergy might be in the provisions within the DSA in respect of recommender systems. This 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. The Irish Safer Internet is aware of the fact that this area requires more research (Irish Safer Internet).

Similarly, the WeProtect Global Alliance position is that 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 (WeProtect Global Alliance).

Eurochild believes 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 instance, the Code could include provisions to carry out child rights impact assessments when developing recommender systems or algorithms. The Code should be ambitious and effectively enforce and extend some provisions from the DSA, where possible. They propose to extend 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 for the DSA Coordinator (Eurochild). According to the Rape Crisis Network Ireland, they 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 (RCNI).

Women's Aid agrees the Code should maximise synergies with the DSA. The focus of their response is that the Code should be designed with the objectives: to require commitments (and ensure mechanisms to evaluate) cooperation between platforms to minimise the burn out on a victim/survivor having to deal with multiple platforms relating to a single or connected experience of online harm; and to require all platforms to have - or sign up to - a meaningful commitment to recognise 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 minority backgrounds and circumstances (Women's Aid).

The Dutch Ministry does not have a proposal on how to minimise the potential for conflict and maximising the potential for synergies between the Code and the DSA. However, they are very interested in Ireland's approach and the relationship between its national legislation and the DSA, particularly how national legislation concerning similar topics and the DSA can coexist (Ministry NL).

The German self-regulatory body (FSM) notes the differences between AVMSD and DSA. According to FSM, the AVMS Directive 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 focuses on illegal content. They believe that from a user perspective, it will be difficult to determine on which legal grounds certain content is inadmissible. That is why when drafting this Online Safety Code, the Commission should have the upcoming execution of the DSA in mind (FSM).

The BFLGI point out that 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. They reiterate the importance of the evidence showing that some major companies are not sufficiently protecting children from online harms (BFLGI).

According to the Irish Heart Foundation, existing legislative provisions are lacking detail in their implementation. 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. 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 (Irish Heart Foundation).

The Internet Watch Foundation points out that 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. They 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 risk 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 (Internet Watch Foundation).

The response to this question from the DCU Anti-Bullying Centre indicated a certain uncertainty as to where the Code under the Online Safety and Media Regulation Act might conflict with the DSA. Namely the DSA contains provisions for voluntary codes of conduct only (Recital 103 p. 28, Articles 45, 46), whereas the codes under the Act are binding for the designated platforms (DCU- ABC).

Spunout believes it is imperative that where any conflict exists 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, Spunout emphasises 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 (Spunout).

The industry has a different view on the future Code. In principle they warn against going beyond the measures required by the AVMS Directive as this approach would risk 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'. They emphasise that 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. 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 Online Safety and Media Regulation Act (Google).

Hence, the Code should be principle based. This means moving away from reliance on detailed, prescriptive rules. More detailed provisions could be included within non-binding guidance, 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. In the view of industry, the non-binding guidance can be updated more easily (where required) to ensure consistency against such codes as they emerge (Google).

The industry shares particular concerns about the conflict and overlap between the AVMS Directive, Online Safety and Media Regulation Act and the DSA in relation to the following areas: (i) Content limitation orders: 'Legal but harmful' content, (ii) Trusted flaggers/nominated bodies, (iii) Complaints-

handling and dispute settlement, (vi) Transparency reporting, (v) Risk assessments and audits, (vi) Blocking orders, (vii) Fines and penalties and (viii) Criminal sanctions as an enforcement mechanism.

When it comes to the "Legal but harmful" content, the industry points out the results of the EU Stakeholder consultation on the DSA proposal where there was a general agreement among stakeholders that 'harmful' (yet not, or at least not necessarily, illegal) content should not be defined in the Digital Services Act and 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). The DSA only allows Member State authorities to order the removal of illegal content. This is an instance where conflicts between the DSA and the Code may emerge. 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 (Google).

The industry also highlights the other potential conflicts between the DSA and the Code namely "Member States should not adopt or maintain additional national requirements relating to the matters falling within the scope of this Regulation, unless explicitly provided for in this Regulation, since this would affect the direct and uniform application of the fully harmonised rules applicable to providers of intermediary services in accordance with the objectives of this Regulation". (Recital 9 DSA). They argue that 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 (TikTok).

Furthermore, 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) (Meta).

They also share the opinion that the DSA addresses many of the same issues as AVMSD/OSMR and the Commission 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 the 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 (Meta).

The industry additionally points out that 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 the Commission 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, the

Commission should take into account that the EU legislature chose to exempt non-VLOPs from those obligations (Meta).

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.

(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.

(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.

(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).

(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.

(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".

(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 (Meta).

Regarding harmful content, the industry shares the opinion that 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 (Meta).

The industry also calls for further harmonisation with other EU content regulation which will be of key importance. The Code should ensure 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 (TikTok).

4.4 How to address the 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?

Regarding content that is connected to video content such as comments posted by users who have viewed videos, descriptions of videos or text and images embedded within videos, which can change the impact the videos have, the following section summarises the responses of stakeholders.

There are two main arguments related to this issue: one argument is that this additional content should be included in the Code as such content can enhance harm or introduce harms or illegal behaviour to audiovisual content; and the second argument is that extending the four areas addressed by the AVMS Directive measures to ancillary features like comments would be disproportionate and go beyond its intended scope.

The key areas of concerns that were raised by those who believe that comments related to video content should be included were the following: sexualisation of images and videos of children via commentary; the promotion of hate speech and derogatory statements in comments concerning certain groups in society featured in videos; bullying; content linked to the intimate images shared without consent, including derogatory, offensive, threatening and abusive comments; other content linked to the intimate images shared without consent, including names and contact information and home addresses of the victims of the sharing of images; homophobic, abusive and threatening comments; links to harmful content and deepfake videos.

The Irish Safer Internet Centre carried out a partner survey with the National Parents Council (NPC), for the purpose of the Call for inputs. This survey aimed at both parents and children and young people, revealed that: “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, NPC response, Irish Safer Internet Centre).

The Department of Health expressed the opinion that in order to be truly effective, 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 (Dept of Health).

A holistic approach was recommended to include such content in the Code as "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" (Brian O' Neill).

The opinion of Safe Ireland 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. They 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 (Safe Ireland).

As noted earlier (see above), Eurochild states that 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.) (Eurochild). The issue of links to external content was also raised. 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 (SCU/SLU).

The Irish Safer Internet Centre stresses the importance of dealing with images and videos which may not in themselves be illegal, but the commentary added is a reflection of, and an encouragement of illegal behaviour. The Centre elaborates on examples from research - The I-KiZ – Centre for Child Protection on the Internet Paper "Combat of the Grey Areas of Child Sexual Exploitation on the Internet". This research shows that innocent images and videos can become exploitative with the addition of commentary with the purpose of sexualising children. In particular the submission cites examples of children in swimwear at beaches to which a range of highly exploitative comments were added. The activities of users and group members commenting act as reinforcement. Many users in relevant groups used depictions of children as profile images, which in this real-life context can indicate a sexual interest in children. In the research, the screening of likes, comments, friend lists and memberships of other groups also 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 acted upon 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 problem can apply to other types of content, such as cyberbullying which is heavily context dependent (Irish Safer Internet Centre).

The Children's Rights Alliance members have stated 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 harmful but when it is considered alongside the content, such as comments connected to the video, it can cause significant distress and harm. Members of the organisation have reported 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. Hence, the Alliance recommends that 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 (Children's Rights Alliance).

According to Belong To, in 2023, they released findings relating to the experiences of LGBTQ+ young people living in Ireland and their social media use. A shocking 87% of LGBTQ+ youth had seen or experienced anti-LGBTQ+ hate and harassment on social media in the previous 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 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. Community guidelines emerged 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 (Belong To).

Examples were provided of where content linked to intimate images shared without consent could be extremely harmful: for example it can identify or locate the person that falsely suggests the person provides sexual services; there are examples of posting of videos on escorts sites, without the woman's consent or knowledge and including their phone number, social media profiles or address; 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).

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. They also recommend that in any case abusive, misogynistic and violent comments should not be allowed and platforms should be required to develop policies recognising gendered violence and abuse and set out both their commitments to eliminating this and tangible actions to address this in the round on their platforms (Women's Aid). According to the RCNI, 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 to the video content itself (Rape Crisis Network Ireland).

Regarding cyber-bullying, it is noted that this takes place in conjunction with comments, videos, and images, rather than being limited to videos alone. Where certain videos may not be harmful on their own 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. The DCU Anti-Bullying Centre cited research that showed that 37-59% 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. DCU ABC believe that 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 should take initiative to educate their users about the importance of thoughtful and considerate captioning. All of the above should take into account the added complexities of live-streaming audio-visual content which, according to the most recent evidence, may necessitate its own legal framework (DCU-ABC). Live-streaming audio-visual content was also referenced as a particular threat in relation to certain types of content such as child sexual abuse (WeProtect Global Alliance) and suicide (National Suicide Research Foundation - UCC).

The DCU submission also discusses Generative Artificial Intelligence (AI) tools, 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 with a new avenue for audio-visual creation. While noting that the effects of such fabricated audiovisual content on propagation of cyberbullying and other online harms are yet to be fully grasped, they point to reports that suggest this is of growing concern. To that effect, VSPSs providers should incorporate distinct markers or labels 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 (DCU-ABC). The response also references the European Union Artificial Intelligence Act with reference to requirements regarding the labelling of AI generated content.

In relation to commercial communications, and to the extent that this could happen, the Advertising Standards Authority of Ireland (ASAI) 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 (ASAI).

The Dutch Media Regulator (CvdM) stated that they believe that it is 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. There is often an entire community (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. Hence, they state that 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 (CvdM). The Dutch co-regulatory body recommends an automated analysis of connected contextual content (NICAM).

Some industry responses emphasise that the AVMS Directive specifies that the “appropriate measures” VSPS should take to protect users apply to “programmes, user-generated videos and audiovisual commercial communications”, as opposed to purely ancillary features, such as comments. Despite this some platforms already address these issues. For example, it is noted that YouTube takes issues with content connected to video content seriously. Between January and March 2023, YouTube blocked over 850 million comments, detected through a mix of automated and human flagging. From the experience of Google, 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. In relation to this, industry responses emphasise 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 would risk cutting across the harmonised approach required by the DSA (Google).

It was noted in the response from UCD that YouTube had turned off comments on videos that feature children in 2019 due to the prevalence of predatory comments. While, for children's accounts for those under 13 comments are not available, children between 13 and 18 remain unprotected. They 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, 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 (UCD).

There is a further nuance in the discussion to be noted in some responses as opinions differ as regards the potential approach to addressing harmful comments – and whether they fall under the rules for removal or the approaches to risk mitigation. Under the section focused on the priorities of the Code (above) where the scope was addressed, some industry responses emphasised that the Code should not go beyond four key areas of AVMS. Going beyond this would include extending the measures to ancillary features. 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 (Google).

According to others, 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 they 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, it is emphasised that 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 (Meta).

Likewise, the same opinion is expressed by TikTok that 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 VSPS. To ensure legal clarity and to avoid potential patchwork implementation across the EU, the Code should avoid deviating from the requirements of the AVMSD.

Hence, they 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 (TikTok).

WeProtect Global Alliance have carried out research into link-sharing and child sexual abuse that 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. The response also described feedback about the value of 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 (WeProtect Global Alliance).

5 Measures to be taken by Video-sharing Platforms

5.1 Online safety features for users

This section provides an overview of the responses to questions 8, 9, 10, 11, 12 and 13.

5.1.1 Feature for Declaring Commercial Communications – Measure (c)

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 Competition and Consumer Protection Commission stressed the importance of users of VSPS being made aware of 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 provisions to protect them against misleading advertising or marketing with the potential to create consumer detriment (CCPC).

Responses that focused on harmful commercial communications and children were discussed above. Some respondents stressed the significance of the child's limited understanding of commercial communications such as: limited ability to understand the persuasive intent (under 8 years) and a lack of abstract thinking skills that help children recognise advertising as a larger commercial concept (8-11 years). Therefore, they reference the Council of Europe recommendation 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' (Children's Rights Alliance, Irish Heart Foundation, Eurochild, DCU-ABC).

It is proposed that the Code should 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. This means it should make sure the child, regardless of their age, understand what "contain commercial communications" means and is able to answer meaningfully (Children's Rights Alliance, Eurochild, Irish Safer Internet Centre).

The National Parents Council states that 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 (NPC).

The Irish Heart Foundation cite the UN Committee on the Rights of the Child who 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." In addition, the Committee have

recommended that there is a need for codes 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' (IHF).

SCU/SLU is of the opinion that content creators should have to declare the presence of commercial communications 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 the 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 (SCU/SLU). Regarding the visibility of banners, the Dutch Commissariaat de Media shared their experience of supervising such content where the supervisory team are unable to find the declaration, either because it is so small or because it is in an indistinct colour. In some instances, the format contained a white font which is not sufficiently visible in a video with a white background. Another issue was that after accepting the cookie policy whilst watching the video, the declaration may not show up when watching the video again (CvdM).

It is also stressed that the Code should also take into account the reliance on child models and child promoters on influencer platforms. Guidance around this should be issued 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 (Cybersafe Kids). 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' (Dutch Ministry). The Dutch regulator states that they have noticed that video uploaders often use multiple hashtags with only the final one containing the declaration of, for example, advertising. The Dutch Commissariaat strongly believes that such a declaration should be the first hashtag (CvdM).

The Advertising Standards Authority of Ireland proposes that the Code be based on high level principles supplemented with guidance notes. A good example is the ASAI code which requires that all advertising must be designed and presented in such a way that it is clear that it is a marketing communication. For influencer marketing they have developed guidance and are currently developing joint guidance with the CCPC (Competition and Consumer Protection Commission). 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. They also propose specific reference to existing guidance rather than drawing up new detailed rules as it leans into the existing work of the CCPC and the ASAI (Advertising Standards Authority of Ireland).

Research published in 2022 by the CCPC on online consumer behaviour and influencer marketing found that consumers may be overconfident 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 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 were analysed were either not labelled at all or not sufficiently labelled. The CCPC engaged directly with consumers and influencers and found that there was widespread agreement amongst both groups that clear guidance would be beneficial for everyone. Hence, they believe 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. Examples of this include in Denmark where 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 (CCPC). A similar opinion regarding straightforward tags such as #advertisement or #paidpromotion is shared by Spunout and the Dutch Ministry.

Another vitally important issue is transparency, 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. According to the Dutch regulator, 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 would increase the transparency of their identities (CvdM).

The industry acknowledges the need 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 is also in line with the AVMS Directive and the DSA (Article 26(2)). TikTok reiterates that a reasonable and proportionate approach 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 (Google). They 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 (TikTok).

Google provides 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. Google developed policies considering to be best practices. These practices include: (i) Clear policies regarding paid product promotions, sponsorships, and endorsements, (ii) Automatic disclosure messages overlaid at the beginning of videos and (iii) Encouraging uploaders to understand the laws and regulations around paid promotion in their jurisdictions (Google).

Google, Meta and TikTok 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. The principles-based approach 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 (Meta). Meta further points out that 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 (Google).

As an example, on YouTube, they 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. They 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 (Google).

Another example concerns various measures already adopted by Meta Platform Ireland Limited (MPIL). They believe those constitute best practice, given the prescriptive nature of the relevant provisions and the intent of the DSA. Some of the measures are summarised as; (i) Advertising transparency: As part of MPIL's compliance with Article 26 of the DSA, users are able to see who is benefiting 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, (ii) 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 repository (Meta).

TikTok also has a number of tools in place to ensure that advertisements and other commercial communications are clearly and effectively labelled for users. TikTok state that they are transparent with users about its approach, and they have included links to the simple, concise guides they make available to users: (i) 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, (ii) 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") (TikTok).

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. 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 (Irish Safer Internet Centre, National Parents Council).

5.1.2 Flagging Mechanism – Measures (d) & (e)

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?

How to ask VSPS providers to introduce and design a flagging mechanism in the Code

As far as the design of flagging mechanism is concerned the Children's Rights Alliance and the Irish Heart Foundation point out that 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 made a reference to the Council of Europe (COE) 'Guidelines to Respect, Protect and Fulfil the Rights of the Child in the Digital Environment'. The COE 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'. As an example of how to design a flagging mechanism that responds to the rights of children and young people, they reference the UK Children's Code regarding the protection of children's data online. This model could be taken and adapted to specifically relate to video content for the purposes of the Online Safety Code. They recommend the best interest of the child should be a key focus when considering the design of the flagging mechanism. 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 (Children's Rights Alliance, Irish Heart Foundation). A similar opinion was expressed by the DCU-ABC.

The Irish Safer Internet Centre proposes that key features of a flagging mechanism should be designed based on the principle of minimum standardisation of fundamental aspects that are universally applicable to all VSPS. They suggest following the best practice examples such as YouTube's priority flagger program (Irish Safer Internet Centre). Safe Ireland's view is that the flagging mechanisms provided by VSPS 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).

The Dutch Ministry recommends the consideration of 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 (Ministry Netherlands).

Belong To is of opinion that while a user flagging mechanism is important, it should not be a primary means relied upon to address harmful content. Firstly, through the Code, VSPS and other social media sites should be bound by a duty of care towards their users, meaning that the onus should be on platforms to address this harmful content before it reaches a critical mass of users. Secondly, 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 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 (Belong To).

WeProtect notes the requirements under the DSA 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. They add an option for users to be able to contest decisions by VSPS if content flagged by them as inappropriate is not actioned. They also point out that caution should be taken that not too much onus is placed on users to report (WeProtect Global Alliance).

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. They additionally stress the following: (i) Consideration should be given to the evidence of the effectiveness and dependability of generalised 'trigger warnings' (ii) 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 and routinely reviewed and validated with relevant support services and accurate at all times. (iii) 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 international. 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 (HSE NOSP).

Regarding the design of a flagging mechanism in the Code, reference was made to the emerging international consensus on standards that should apply in the design of online safety features that conform to principles of Safety by Design (SbD). 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. 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. This includes tools aimed at companies for assessing how systems, processes and practices support user safety based on principles and good practice in SbD (Brian O'Neill).

The 5Rights Foundation works with the Institute of Electrical and Electronics Engineers (IEEE) to create a suite of standards for Age-Appropriate Digital Services. They will be based on the 5Rights principles for children, which include: (i) presenting information in an age-appropriate way; (ii) upholding children's rights; (iii)offering fair terms for children; (iv) recognising childhood; (v) and putting the child ahead of commercial interests and ahead of platform status (Brian O'Neill).

How to ensure that VSPS providers introduce the mechanism in a user-friendly and transparent way

The Irish Safer Internet Centre suggest that a user-friendly flagging mechanism should be provided 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. There is a need for accessibility by design on the platform (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 timeframes (Irish Safer Internet Centre).

In support of better transparency, the SCU/ SLU recommends that VSPS providers engage an Independent Moderator and seek the input of a professional with child protection expertise. The Independent Moderator would be connected with the Commission. The findings of the audit should then be published, with recommendations made. This process should be underpinned by statutory powers (SCU /SLU). A similar opinion is shared also by Belong To.

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 (Women's Aid, Safe Ireland).

The Children's Rights Alliance describes the approach in the UK, where the ICO's (Information Commissioner's Office) has created guidance for services for 'age-appropriate design.' This states that 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. 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. 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. The ICO provide examples of how to do so in the Code for each age range from 0- 5 up to 16-17. 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. 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. Additionally, in order to 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. This model could be taken and adapted to specifically relate to video content for the purposes of the Online Safety Code (Children's Rights Alliance).

The Dutch regulator (CvdM) proposes that '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. 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'. The Dutch Ministry is in also examining 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 (CvdM).

The German self-regulator FSM believes that 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 (FSM).

The industry 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 (Google).

Many providers, including Facebook and Instagram, already have 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. They have made the reporting option even more user-friendly as they work with content designers and user experience experts to ensure that such mechanisms are accessible and easy to use and understand (Meta). They 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 (TikTok).

How the Commission should ask VSP Providers to report the decisions made on content after it has been flagged

Eurochild proposes that some common minimum standards should be established as evidence shows that children often tend to block content instead of reporting when they feel at risk. This is due to 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 categorise 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. Also, the system 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) (Eurochild).

The recent National Parent Council Survey indicates that "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).

The Irish Safer Internet Centre recommend that the flagging mechanism 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 (Irish Safer Internet Centre).

Women's Aid shared their view on reporting mechanisms, detailing what can be expected by the VSPS after a report is made and within which timeframes. They 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 Image-Based Sexual Abuse 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. Women's Aid conclude their proposal by stressing the importance of making the process for flagging content as straightforward and easy to understand for children and young people as possible (Women's Aid).

Safe Ireland proposes that real-time data on numbers of notifications received, those being worked on, those 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 (Safe Ireland).

The Dutch regulator (CvdM) is of the 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 (CvdM).

Regarding reporting the decisions made on content after it has been flagged, the German self-regulator FSM states that users will want to know if their report was taken care of so the provider should always send an appropriate response, preferably 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 the 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 (FSM).

The Dutch Co-regulator NICAM outlines their experience with reporting decisions after flagging. This has shown that an independent organisation 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 (a system of icons used to classify content in the Netherlands) they 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 their website (NICAM).

The Department of Health proposes the inclusion of harmful content related to suicide, self-harm and eating disorders in the flagging mechanism. This feature should be accompanied with information on appropriate supports and services. 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) (Department of Health).

The Samaritans emphasise research-based evidence, alongside the experience of the Samaritans Central Charity advisory service engagement, which indicates that 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 study determined the half-life (or lifespan) of social media posts on different platforms: Snapchat (0 min), Twitter (24 min), Facebook (105 min), Instagram (20 h), LinkedIn (24 h), YouTube (8.8 d), and Pinterest (3.75 months). 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. They also emphasise that 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 (Samaritans).

Regarding the question about reporting the decisions made on content after it has been flagged, the industry draws attention to the fact that turnaround times for illegal content is already harmonised by the 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, Meta believes that it would be inappropriate for a sectoral regulator to seek to specify timing by means of an online safety code (Meta).

To assist the Commission in determining what may constitute best practice for reporting decisions on content after it has been flagged, TikTok provided information on their existing approach which supports reporting of content made both under their community guidelines and illegal content under DSA. In keeping with their commitment to ensuring procedural fairness, they seek to provide notifications to community members if they have violated the rules. If a user posts content that is not allowed 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 (TikTok).

The extent to which the Commission should align the Code with similar provisions on flagging in the DSA

The general view on this question is that the Code should be aligned with the DSA. The Children's Rights Alliance propose that the DSA (Article 16) should be integrated into the Code. Article 16 of the DSA 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 (Belong To, Children's' Rights Alliance, Safe Ireland, Women's Aid). 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. 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 (Children's' Rights Alliance, Belong To, Safe Ireland). Likewise, it is recommended to have full alignment where possible (Irish Safer Internet Centre, FSM).

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 (Department of Health, Safe Ireland). Regulatory alignment with other regimes is important and the Internet Watch Foundation favour alignment with Article 16 of the Digital Services Act which sets out criteria for trusted flagger programmes in respect of illegal content (Internet Watch Foundation).

The Code should align with similar provisions on flagging in the DSA. 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. 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 Google would encourage the Commission 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 (Google).

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 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 (Meta). TikTok emphasises 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 (TikTok).

5.1.3 Age Verification and Age Assurance Features- Measure (f)

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?

5.1.3.1 Question 10a: What requirements should the Code include about age verification and age assurance?

As regards requirements that should the Code include about age verification and age assurance, responses were extensive and detailed and are outlined below.

The 5Rights Foundation proposes 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 (5Rights).

The following common approaches to age assurance were cited also by a range of respondents. Depending on the purpose, context and level of risk, services may implement a combination of approaches to age assurance. A lack of minimum standards may lead to the exacerbation of known problems of excessive data collection, privacy infringements, ineffective age checks and could lead to heavy-handed age-gating that can block children out of spaces they have a right to be in (Children's Rights Alliance, Irish Safer Internet Centre and Ombudsman for Children). 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 these principles:

- 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.

In addition, 5Rights emphasises that 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. 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 (5Rights).

Reference is also made to the Irish Data Protection Commission's 'Fundamentals for a Child-Oriented Approach to Data Processing', which 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 (Ombudsman for Children). 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.

The Children's Rights Alliance adds that 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.' 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'. As an example, they refer the UK Children's Code regarding the protection of children's data online, which offers a dual option to designated services in order to comply with the standard of age verification or 'age-appropriate application'. (i) 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 (ii) by applying the standards in the Code to all users. The Children's Rights Alliance further recommends that:

- 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 complaint 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 (Children's Rights Alliance, Belong To).

In the opinion of the Dutch regulator, self-declaration is not an appropriate age-verification tool, insofar as it does not actually verify someone's age and is easily to circumvent. 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) (Commission for the protection of Minors), is an age verification system like Yoti (CvdM).

The Department of Children, Equality, Disability, Integration and Youth raises another concern regarding robust age verification systems such as the need to provide identity documents or use a proposed European Digital Identity. They claim this has the potential for digital exclusion of young people from marginalised groups. Some vulnerable groups in society may be less likely to have identity documents, and it will be important to study what effect such a system would have on digital access for young people who are of age, 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 (DCEDIY).

Belong To raises an additional consideration to be taken in the case of LGBTQ+ young people. An international research study 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. 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. 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 (Belong To).

According to Samaritans Ireland, 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 (Samaritans).

WeProtect Global Alliance believes that age assurance is one of the tools that can be used to create digital products safe by design. In their 2021 Global Threat Assessment they 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 (WeProtect).

The DCU-ABC reports that in light of significant underage use of social media in Ireland (they cite the NACOS report, 2021), it is 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 developing technologies that are being used for such purposes, it is still difficult to understand how specifically VSPS engage 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.

DCU-ABC understand the privacy and freedom of expression-related concerns when mandating document-based age-verification from all users and would not support those. Advancements in biometric age verification in terms of data minimisation and minimising potential for privacy infringements appear promising, but they have not conducted research themselves in this domain. Transparency in terms of effectiveness of age-assurance and verification procedures appears to be lacking. DCU-ABC 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 they 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? They were not able to receive answers to these questions, which appear to be very simple compliance-related questions (DCU ABC).

In the opinion of the DCU-ABC view this is a critical issue as 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 disincentives companies to innovate for underage users. Furthermore, they 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 (DCU ABC).

Spunout focuses on the Digital Age of Consent (DAOC)', which they cautioned against the Government's setting DAOC at 16. Their 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. Their 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, the 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, platforms should never be absolved of their responsibility to keep their products safe for children (Spunout).

Spunout believes that the 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. The current approach arguably risks incentivising children to lie about their age to get online. This seriously undermines the ability to make online and video-sharing spaces safer for children, as it means it won't be possible to know how many children are actually online (Spunout).

Regardless of the above, the Digital Age of Consent is currently set at 16 and Spunout accept that this is unlikely to change in the immediate future. However, they 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 the problem is moved 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. Spunout states that they are 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 (Spunout).

The Age Verification Providers Association (AVPA) supports a proportionate approach to age assurance, with the Commission stipulating the level of age assurance required for the most commons use-cases, based on the latest international standards. They believe it would be 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. AVPA warns the Commission not to assume that the eIDAS Wallet will solve the challenge of online age assurance. They 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. 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 (AVPA).

'Verify my' advocates the use of methodologies independently certified as meeting the requirements of PAS 1296:2018 – Code of Practice for Online Age Verification. They support re-usability, such as their VerifyMyAge integrated platform which allows the user to stay verified across any VerifyMyAge integrated platform. This platform is based on using ID scan, credit bureau check, mobile phone number check, credit card check, and provides an estimate of a user's age or age range relying on artificial intelligence and machine learning techniques e.g., facial age estimation, voice age estimation, email address check (Verify my).

The details of age verification systems are not specified in the AVMS Directive, and it is the case that these 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. It is for industry providers to demonstrate that adequate safeguards are in place to ensure that content that "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 (Brian O'Neill).

Google's position is that 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, they state that they 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". They are continuously looking at ways to best create an appropriate environment for family content on YouTube, so they 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, they 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 (Google).

Google illustrated the point that industry has already come up with innovative, effective and proportionate solutions, in the absence of prescriptive regulation. For example, they outlined the age assurance and age-verification methods currently employed by YouTube

They 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 their systems are unable to establish that the user is above the age of 18, they will request that they provide a valid ID or credit card to verify their age. They have built the age-verification process in keeping with Google's Privacy and Security Principles.

Their 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. They do not collect new information from users in order to run this age inference model (Google).

Meta shares a similar position to that of Google. They believe that the Commission should take into account the Data Protection Commission's Fundamentals for a Child Orientated Approach to Data Protection (the Fundamentals), which outlines various recommendations relating to age verification, which already apply to many (if not all) VSPS. Likewise, the Commission 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, the Commission should consider whether to address this topic at this time and, in case it decides to do so, the Commission should adopt a principles-based approach that is consistent with other EU age assurance efforts. Meta also provides some details of the core practices which they 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. They have found that encouraging and facilitating processes for easy reporting of underage users is a proportionate measure, since they 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, they consider it best practice to enable simultaneous disabling across services where a user has been flagged as underage.

(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 (Meta).

TikTok states they are also deeply committed to the safety and privacy of their users, especially their younger users. In particular, they are committed to preventing under 13's accessing their platform and to continuing to enforce their policy on the platform by detecting and removing younger users who are not old enough to access the platform. Preventing underage people from creating an account:

- TikTok has a 12+ rating in the App Store and GooglePlay, which enables parents to use device-level controls to block their teens from downloading TikTok.

- To help keep people from using the platform if they're not old enough, they have 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 the minimum age requirement, they suspend their ability to attempt to create another account using a different date of birth.

Regarding removing suspected underage accounts:

- The commitment to enforcing minimum age requirements does not end at the age gate, and they take a number of additional approaches to identify and remove suspected underage account holders.
- They train the safety moderation team to be alert to signs that an account may belong to someone under the age of 13. They use other information provided by users, such as keywords and in-app reports from the community, to help find potential underage accounts.
- When the 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 a moderator 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 the underage moderation team.
- To bring more visibility to the actions taken to protect minors, they are the only major platform to regularly disclose the number of accounts removed from the full TikTok experience for potentially belonging to an underage person.

TikTok state that they are committed to exploring innovative solutions in these areas. They believe the industry should work toward accessible, robust, privacy preserving options. Currently, there is no "silver bullet" age assurance solution that can be 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, they would be keen to explore it further. In the meantime, they are working towards further enhancing their age assurance strategy. They 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. They 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; and the EC's Strategy for Kids (TikTok).

Summary of the Irish Data Protection Commission's "Fundamentals for a Child Orientated Approach to Data Protection"

The guidance regarding age verification is structured in 5 chapters. In the First Chapter (Age of Digital Consent) the concept of the so-called "age of digital consent" in relation to information society services stems from Article 8 of the GDPR, which states that if an information society service (such as a website or an app that offers a service, e.g. gaming, social media, video-sharing, etc.) is being offered directly to a child, and that service is relying on consent as the legal basis to process the child user's personal data, then parental consent (described in the GDPR as consent from the holder of parental responsibility) must be obtained (in Ireland where the child is under 16 years of age). Of critical importance is the fact that the requirements around the age of digital consent do not impose restrictions on a child being able to access a service. The age of digital consent is also a marker for

online services to consider the nature and design of their services, and how to make them age appropriate for their users. Digital consent obtained from children over the age of digital consent (i.e., 16 or over in Ireland), or from the guardians/parents of children under the age of digital consent, should not be used as a route to treat children of all ages as if they were adults.

The Second Chapter (Verification of Parental Consent) addresses verification of parental consent. Given the scale of technical specialities and resources available to technology and internet companies (i.e. whose business models are predicated on deployment of digital and online technologies) and the higher risks to the data protection rights of users who utilise their services, especially children, the DPC considers that a higher burden applies to such organisations in their efforts to both verify age (see below) and verify that consent has been given by the parent/ guardian of the child user.

In the Third Chapter (Age Verification Purposes) the terms "Age verifications" vs. "Age Assurance" are outlined. Age verification is 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 child with a high degree of certainty (e.g., government–issued ID, electronic identification services, secure third-party services, etc.). Age assurance on the other hand, is a broader term given to the wider range of methods that can be used in addition to age verification to estimate the age of a child user. An age assurance approach allows organisations to select methods that are most suited to the specific risks involved in their processing. Age verification when undertaken by organisations in support of reliance on consent under Article 8 of GDPR as the legal basis for the processing of personal data is just one of the situations in which age verification methods may be employed. The DPC recognises that there are other purposes for age verification undertaken by an organisation, including:

- allowing access to its service – for example where an organisation provides an adult-only service, which by law it cannot provide to under 18s e.g., gambling related services; and
- providing a "child-friendly" version of a service which attracts a mixed user audience i.e., by offering enhanced data protection settings/ features for child users, in line with the requirements of these Fundamentals, particularly the Floor of Protection.

The Fourth Chapter is focused on age verification and the child's user experience. While an organisation's choice as to whom it will offer its services falls outside the scope of data protection per se, the DPC's position is that the specific requirements of Article 8 (including the associated implication that age verification underpins verification of parental consent), or any other obligation under the GDPR, including compliance with the requirements of these Fundamentals, in no way justify the "locking out" of children from a rich user experience simply on the basis of purported data protection compliance. Chapter Five is about minimum user ages. The DPC notes that it is common practice amongst many of the most popular online service providers to apply a minimum user age of 13. However, the DPC does not consider that the setting of a minimum user age obviates the requirement for such service providers to comply with their obligations towards child users below this age, where children are likely to use the service in question. Where a service provider stipulates that their service is not for the use of children below a certain age, they should take steps to ensure that their age verification mechanisms are effective at preventing children below that age from accessing their service. If the organisation considers that it cannot prevent children below its stipulated age threshold from accessing its service, then the organisation should ensure that appropriate standards of data protection measures are in place to safeguard the position of child users, both below and above the organisation's official user age threshold. Chapter Six is dealing with age verification mechanisms. There is no one-size-fits-all solution to the issue of age verification. Appropriate age verification mechanisms

are likely to vary from context to context, depending on, for example, factors such as the service being provided, and the sensitivity of the personal data being processed. In any event, such measures 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 processing of personal data undertaken by the organisation is of higher risk to the user based on the criteria identified below in Chapter 7. In Chapter Seven a number of criteria for a risk-based approach to age verification is listed. If organisations decide to implement age verification mechanisms, there are certain minimum criteria which should be considered when determining the approach. What may be considered a suitable approach for one organisation may be entirely unsuitable for another. The following list contains a non-exhaustive selection of criteria which should be taken into account in adopting a risk-based approach to verification:

- 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, hobbies or interests, etc.
- The sensitivity of said personal data - e.g., special category personal data, or data which could be considered sensitive such as financial information, information on family circumstances or birth status or data incorporating third party data 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 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/make available personal data, or elements, 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 (Data Protection Commission - "Fundamentals for a Child Orientated Approach to Data Protection").

5.1.3.2 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?

The Irish Safer Internet refers to the NPC's survey where parents were asked what types of age ratings (if any) should be applied for different video content. 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 an Age Gating requirement. 33% of the young people who responded to the NPC survey felt it should be via 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 (NPC Survey 2023, NPC, Irish Safer Internet).

The Dutch regulator expresses the opinion that only content that the VSPs can determine with certainty as not being harmful to the physical, mental, and moral development of minors, should be shown by default to users who are logged out or in private browsing mode and whose age cannot be verified or assured. A good example of how to set this up is the way that Ofcom did in their VSP regulation, which is to divide VSPS into segments/risk groups (for example 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 VSPS, type of audience (CvdM).

NICAM strongly supports the view of VSPS being, in principle, 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 (NICAM).

AVPA's opinion is that this question confines itself to content, but of course online harms arise also from contact, conduct and contract (the 4 C's). 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 (AVPA).

Regarding the sort of content shown by default "Verify my" proposes a reliance on the Irish "Fundamentals" Code, which applies largely to any business or "information society service 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 (Verify my).

5.1.3.3 Question 10c: What evidence is there about the effectiveness of age estimation techniques?

Several responses reference the euCONSENT project, the eIDAS credentials EU Digital Identity Wallet as well as the current works of the Age Verification Providers Association on the BSI PAS 1296:2018 (Eurochild, Irish Safer Internet Centre and the Ombudsman for Children).

According to their responses, the European Digital Identity Wallets do 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. The euConsent system aims to deliver a pan-European, open-system, secure and certified interoperable age-verification and parental consent to access Information Society Services. From their experience of working with children, Eurochild found that their level of trust towards ID-based age verification systems is alarmingly low. From this, the Eurochild believes 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 (minimising the personal data collected and stored by platforms when verifying age, i.e., name, address, or date of birth) (Eurochild).

The same opinion is shared by the Irish Safer Internet Centre, which in addition recognises that that the proposed European Commission funded euConsent project will not be live for another 12-18 months. The euConsent project is 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 of the project 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 (Irish Safer Internet Centre).

The Irish Safer Internet Centre references Yoti's white paper on the effectiveness of its age estimation techniques. It reports high true positive rates, including for minors. YOTI is a recognised model used by child protection agencies such as the NSPCC (National Society for the Prevention of Cruelty to Children) and IWF. It is emphasised that reference to industry products does not constitute an endorsement on behalf of the Irish Safer Internet Centre (Irish Safer Internet Centre).

The SCU/SLU opinion is that 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. They believe 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 parental 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 (SCU/ SLU).

The German self-regulator FSM has recently seen an enormous development in the effectiveness of age estimation techniques. FSM member Yoti continue to be very vocal on their numbers according. 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 accepted a tool for preventing the access to adult pornography by minors which did not re-quire 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 (FSM).

AVPA points to the selection of evidence from some of their members already in the public domain as indicative of the latest levels of accuracy for facial image and voiceprint analysis. VerifyMyAge offers 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 (AVPA). "Verify my" emphasises their exhaustive choice of available age verification and estimation methods and datasets. According to their research and their own data they reached the point where 99.9% of their verifications of adults are successfully approved. They use several methods such as: Database Check, Mobile Phone Number, Facial Biometrics, Government Issued ID, Credit Card, Open Banking and Email Address (Verify my).

5.1.3.4 Question 10d: What current practices do you regard as best practice?

The AVPA states that it is not arguing for wholly independent age assurance carried out only by third parties. The Commission should consider both the question of confidentiality and expertise. (i) 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. (ii) 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 (AVPA). "Verify my" proposes best practice based on an ecosystem offering solutions across a range of technologies joining efforts with partners such as: Experian, Facetec, Yes, Schufa and Amazon Recognition ("Verify my").

The Trust Alliance Group offers 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, they have identified various practices concerning age assurance across the two cohorts of businesses they assessed. Practices identified during their 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. However, the believe that 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.

They discovered this underlying commercial reality with one of the platforms in their 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, for a short time, to 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 (Trust Alliance Group).

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 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 to be 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 (Trust Alliance Group).

What has become clear through the Trust Alliance Group 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. The Trust Alliance Group opinion on the idea of a child-only platform is that it may raise concerns in what appears to be the ideal location for a predator or bad actor to operate within a walled garden, where they 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. They propose to approach the issue with a sensitivity to the differences between platforms, which 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 (Trust Alliance Group).

5.1.3.5 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?

Where accounts are not age-verified, the view of Safe Ireland 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 (Safe Ireland).

5.1.4 Content Rating Feature- Measure (g)

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?

Many respondents agreed that content rating is important for the protection of minors. The responses described and assessed types of classification and rating systems. In addition, the systems already in place on platforms for content rating were discussed. The importance of such systems being clear and "child-friendly" was also stressed. In relation to "users", distinctions were made between "users" who upload content and "users" who view content. From some of the responses, it is apparent that it was not entirely clear which type of user could be allowed to rate the content – just the uploader, or all users. Some concerns were raised regarding the reliability of such a system and the extent to which content may be rated by bad faith actors.

Others stressed that the existence of a range of other measures to protect children including age verification and parental controls raised questions on the need for content rating. Several respondents emphasised the different cultural contexts of classifying content that may exist in different European countries. It was also recommended by some respondents to move towards a common approach to content rating across Europe.

Data from the National Parents Council 2023 survey provided feedback on the user experience of content ratings and revealed that: 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 (Age-Based Media Reviews for Families | Common Sense Media).

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. Regarding the responses from young people, 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 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% stated that 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 (Irish Safer Internet Centre, NPC survey 2023, NPC Submission).

Content rating is considered to be a broadly welcome feature, whereby it is good to allow parents or guardians to have an insight into the type of content which their child might access. However, it was noted that is it 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 (DECDIY). 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 (FSM). 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 (Brian O' Neill).

Belong To highlight a tendency towards over classification of certain content as being unsuitable for children. Regarding 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 misinformation have warned about the deliberate conflation of age-appropriate information relating to LGBTQ+ people and identities, and accusations of "grooming" and "sexualising" children. 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 (Belong To).

For some, the standards of the Irish Film Classification Office provide a good example of appropriate age rating (SCU/ SLU, IWF). The British Board of Film Classification (BBFC) system is also recommended, and it is suggested that there could be some alignment of approaches between the two organisations (IWF). A further issue emphasised was that 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 (SCU/ SLU). It was also pointed out that the comparison with film classification schemes is not an exact one - as noted by the Australian eSafety Commissioner. In online and user-generated content, schemes such as PEGI and PEGI online may be closer to the VSPS context, particularly regarding the processes followed to rate the content (Brian O' Neill).

The PEGI (Pan European Game Information) was described as being a successful example of self-regulation by providing advice to parents through age classifications. 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. 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 (Irish Safer Internet Centre).

The DCU Anti-bullying centre also discussed PEGI in their submission. They note, however, that 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 the well-being of children without being biased towards the interests of the industry. The recommend instead a system managed by another body even if it collaborates with the industry in order to provide the age recommendations (DCU-ABC).

Finally, they stress that 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) 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. 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 (DCU-ABC).

The UCD submission proposes 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. They provided an overview of methods for auditing content.

Table 1: Adapted from Technical methods for regulatory inspection of algorithmic systems, Ada Lovelace Institute(2021).

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

With regard to the role of "users" rating content, several concerns were highlighted. According to Safe Ireland, this is a very subjective way of determining whether content is suitable, and judgments may 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. They believed 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. A second point questioned (perhaps based on the assumption that various people would rate the same content – i.e., "users" aside from those who upload content) how content ratings of any single video be synthesized. They noted that 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. In addition, Safe Ireland also expressed concern also about content ratings made in bad faith by bad actors with no concern for the welfare of minor users. They did not see how these could be very easily prevented except through extensive and continuous monitoring, which itself is not likely to be workable (Safe Ireland).

A similar concern regarding crowd-sourced "user ratings" was expressed by the industry (Google). Other industry actors expressed the concern that user ratings in general are open to serious abuse and inaccuracies (Meta).

The opinion of the FSM (German self-regulator) was that often users themselves will not be able to provide precise age ratings like those that are well known 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. To deal with these problems, they suggest that 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 (FSM).

In relation to moving towards a common approach to rating, the Dutch media regulator stated that they wished to advocate for a European classification system which applies to all European countries and takes into account cultural differences. They stress that the development of the Code represents an excellent opportunity to move in this direction to harmonise age ratings and content classification in Europe, since VSPs have users spread across the EU. In addition, 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 (CvdM).

The Children's Rights Alliances cites the UN Committee on the Rights of the Child recommendation 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.' They also cite the Council of Europe recommendation 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' (Children's Rights Alliance).

Regarding commercial communications, and whether a 'one size fits all' content rating system would be appropriate, the ASAI noted 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 (ASAI).

Experts reiterated the fact that there are similarities across the many existing rating systems, but there are also many variations due to the many social and cultural differences involved. It was explained that the 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. One of the outcomes of this process was the 'You Rate It system', coordinated by the highly experienced classification bodies NICAM and BBFC, which is 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 (Brian O' Neill).

The NICAM system is further discussed by both the Dutch Regulatory Authority and by NICAM. According to the CvdM 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 NICAM, the Kijkwijzer-system is considered to be valuable by most European parents who also indicated that they understand the system. However, they also note that 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 (CvdM).

Several respondents emphasised the fact that content rating online is complex. It was suggested that 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 (We Protect Global Alliance).

There were several suggestions to maximise efficiency of the content rating systems. According to the Dutch NICAM, research has provided evidence of the types of content that might be harmful to children: 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 member state 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 member state in which certain VSPS are registered, the same up to date criteria and ratings will apply. They propose to join hands on this important topic and spread this message in Europe to make this the success that children deserve (NICAM).

The Dutch regulator recommended 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, in their opinion, 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. This would allow 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’. They recommend including 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 (CvdM).

The number one priority is the content analysis (by the uploaders) of videos, 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 NICAM 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 (NICAM).

The NICAM explain how they further adapted their system to the online world and carried out research on the content available on YouTube, TikTok and Instagram and what elements of this should be deemed potentially harmful. They also asked children about their social media use, experiences with harmful content, and their wants and needs regarding age and content rating. In addition, they also interviewed uploaders, featuring questions such as what is feasible, what technical protection measures and ways of informing the public can be used. Based on this research they developed a special system for these uploaders with which they can rate their own content fast and simply; ‘Kijkwijzer Online for YouTube’. Specific elements like for example ‘dangerous challenges and stunts’ were added to the system. In the analyses they found that these challenges are frequently present on social media and that they can pose risks/ dangers to children. Additionally, they tested the prototype of Kijkwijzer Online among Industry parties on YouTube. The findings with industry so far are 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’ (CvdM).

Therefore, both the CvdM and NICAM requested that the Commission include the obligation in the act for VSPS providers to facilitate (national) rating systems on their platforms by providing their uploaders with options to embed and show ratings on their platform(s) (CvdM, NICAM).

The systems currently implemented by the various platforms were described and discussed. According to the response from DCU-ABC, 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 maturity level. TikTok employs a "Content Levels" system, whereas Twitch employs "Content Classification" system, and YouTube utilises its own content rating mechanism. 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 (DCU-ABC).

From the perspective of the industry, they reiterate the need for 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 (Google). 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. TikTok expresses the view 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 (TikTok).

Google noted that their experience with content rating systems on YouTube has proven effective, although they continue to learn and adapt their 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 (Google).

The issue of the meaning of “users” was also addressed here. Google noted 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. Google combine this with their own classifiers and reviewers to establish the content that should be available only to those over the age of 18. They 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 their experience, subject to abuse (Google).

TikTok has developed Content Levels which organises content based on thematic maturity and giving users choice based on their personal preferences. They emphasise the fact that all content on the platform must comply with their Community Guidelines. Within these strict policies however, they also 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 the 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. TikTok’s approach is similar to what can be seen in the film, television broadcast, and gaming industries. They are drawing closely on the kinds of standards already in use around the world. They have focused on further safeguarding the teen experience first and they plan to add new functionality to provide detailed content filtering options for their entire community so they can enjoy more of what they love (TikTok).

In the opinion of Meta, 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 they also already apply warning screens on graphic content that does not violate their policies for under 18 users. Additionally, like other platforms, they 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 (Meta).

Additionally, Meta believes that such a requirement could potentially create inconsistent experience for users through the European Union. They note that as the Commission 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. In conclusion, they 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 (Meta).

Regarding the overall system of classification of content, 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. They 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 (Google).

Some new challenges regarding content were raised by DCU- ABC. They state that in recent months experts have witnessed the proliferation of Generative Artificial Intelligence (AI) tools that enable users of all age groups to create synthetic content. Additionally, the rise of Deepfake technology, which generates manipulated videos from real footage, poses both positive and negative implications. 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 a functionality to flag videos that are created using DeepFake or Generative AI. For example, videos employing DeepFake to portray another actor discussed in the interview 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. 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 (DCU- ABC).

5.1.5 Parental controls- Measure (h)

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?

5.1.5.1 What requirements should the Code have in relation to parental control features?

The general view of the stakeholders is that parental control is necessary to protect minors using VSPS. Parental control should not be a substitute for the "safety and privacy by design" principle (5Rights, Children's Rights Alliance, Cybersafe Kids, Eurochild, Irish Safer Internet, National Parents Council, Ombudsman, Safe Ireland). Several stakeholders stress the importance of autonomy and child privacy, and impeding children rights online.

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 (5Rights).

The Children's Rights Alliance note that the Council of Europe has recommended that children's evolving capacities should be taken into account when businesses establish or update their parental controls. Additionally, States should ensure that such controls do not reinforce discriminatory attitudes or infringe on children's privacy and information rights. They also provide a reference to the UK Children's Code specifying that if a regulated service provides parental controls, they should give the child age-appropriate information about this. 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. The UK Information Commissioner bases this standard on 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 (Children's Rights Alliance).

Belong To raises a number of additional considerations to be taken into account in the case of LGBTQ+ young people. 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. 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. 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) (Belong To).

Cybersafe Kids provide the results of their research on existing age-restrictions. In short, they believe that age-restrictions do not work. From the 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. 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 these companies point to their parental controls (thereby putting the onus on parents) instead of investing in more substantive child safety infrastructure (Cybersafe Kids).

Eurochild's opinion on parental controls is that they have been proven to be insufficient to ensure child safety online. The Code could be an opportunity to improve some aspects e.g., the interoperability of parental controls among video-sharing platforms. Both the child and the parent should be informed about this by the service to ensure transparency and empower children's agency online (Eurochild).

Findings from the National Parent's Council 2023 survey revealed that the vast majority of parents (95%) were aware or at least somewhat aware of parental controls that are available on digital devices and online platforms, with 49% of them using them regularly, 33% using them occasionally and 17% not 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 (Irish Safer Internet, NPC).

The Ombudsman for Children (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 (Article 5 of the CRC, Article 28(3)(h) of the AVMS). 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 (Ombudsman).

Another proposal is that the Code should contain an obligation for parental controls to exist. This should incorporate features such as, flagging, time limits, content alerts or attempts to change passwords/create new accounts. Parental controls should be turned on automatically (SCU/SLU).

Women's Aid believes that all online platforms should be safe for everyone. Further, they 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. Both, Women's Aid and Safe Ireland recommend that safety and privacy setting for minors should be set at maximum safety and privacy by default (Women's Aid). In addition, Safe Ireland proposes that parental control features should always be visible on any video via an eye-catching, easy to recognise prompt which leads on to simply worded and brief instructions on how to impose these controls (Safe Ireland).

5.1.5.2 Existing examples of best practices in parental control

The Dutch regulator from their experience with international commercial video on-demand services who are based in the Netherlands, such as Disney+ and Netflix, note that these parties have extensive experience with parental control measures. For example, both include 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 (CvdM).

Google's outlines its proactive approach towards online safety as evident in features like YouTube's supervised experience, Family Link and the default settings they have for unverified age accounts. By favouring supervised digital experiences and defaulting to safety-first settings, they ensure an added layer of precaution. They 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. Google work to identify content that is age-appropriate, adheres to 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, they offer a supervised experience. Whichever content settings the parent chooses, the child cannot gain access to 18+, age-restricted content through the account (Google).

Google also expresses the belief that safety is not 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 is paramount for the Code to guide platforms beyond mere control mechanisms, emphasising the significance of user education and awareness (Google).

Based on their experience, the Meta believes that they have struck the right balance by taking an "age appropriate" approach. To this end, they 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. 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; (vi) see which accounts their teen is currently blocking; and be notified if their teen changes any of these settings. They 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 Portal, Parent Centre and Parent's Guide), (2) Family Centre, 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 Hub, 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 (Meta).

TikTok is investing in tools and resources to help parents, guardians, and families support their teens online. They 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. 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. They 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 TikTok's app's Screen Time offerings, including Daily Screen Time and Restricted Mode. Their 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 TikTok considers 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, Filter video keywords, Restricted Mode, Linked Account activity, Search, Discoverability, Suggest account to others, Direct Messages, Linked Videos, Comments, Screen time dashboard and Mute notifications (TikTok).

The WeProtect Global Alliance points out that parental controls and content filters are also key tenets of the Safety by Design approach. In their 2021 Global Threat Assessment, they discovered that 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. The 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 (WeProtect Global Alliance).

5.1.5.3 Should parental control be 'turned-on' by default?

Several submissions provided responses to this question in the context of the section above.

TikTok's opinion is that 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 (TikTok).

By the same token the German self-regulator FSM is of opinion that 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 (FSM). The Dutch Co-regulator NICAM adds that the 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 (NICAM). The ASAI comments they would generally support this approach, as it would be a protective measure for minors (ASAI).

5.1.5.4 Other considerations (Sleep deprivation, Poor mental health, lack of awareness of controls)

The Department of Children, Equality, Disability Integration and Youth states that in addition to blocking harmful content and privacy settings, parental control should include the ability to limit a child's use of a service. It is important to distinguish the issues of children's access to potentially harmful content or 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 in the Irish Times showed correlations between poorer mental health and a lack of sleep and linked the lack of sleep to smartphone use. The Department would welcome parental controls by default and clear and accessible guidance to parents on how to set these controls (DCEDIY).

IWF have seen through the introduction of the Age-Appropriate Design Code several examples of best practice from platforms, in terms of ensuring children's 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. IWF also believes that 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. They support the proposal that children's accounts are set to private by default (IWF).

The Dublin City University Anti-Bullying Centre (DCU ABC) holds a different view on parental control techniques. Their position is that parental control technologies are not universally effective, and they need to be tailored to the evolving capacities of the child. 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. Monitoring and surveillance can also disrupt the parent-child relationship and trust and have a negative impact on a child's right to privacy. The DCU-ABC also stress the need to 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 counter-productive in those cases. They cite research that has found that a positive approach to supervision seems to be more efficient than a restrictive one. 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 (DCU ABC).

Spunout raises the issue of the 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 (from DCU Anti Bullying Centre research). 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. In addition, they emphasise 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 (Spunout). The AVPA aligns with this opinion stating that 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 case as to whether it is more appropriate to require parental controls or online age assurance (AVPA).

Other experts state that video sharing applications should 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 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 (UCD).

Brian O' Neill indicated that parental controls are also a feature of the mobile operators' codes of conduct, as featured in the European Framework for Safer Mobile Use by Younger Teenagers and Children. 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. 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 examines the outcomes for children and families through parental controls as a child protection measure and contains recommendations for design practice. This is also relevant to the design of age assurance systems (Brian O'Neill).

5.1.6 Media Literacy- Measure (j)

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

Respondents addressed the meaning of media literacy, their own work in the area and the proposals for requirements in the Code.

As a useful description of media literacy, the submission of the We Protect Global Alliance explains that as part of their '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 needed to be able to participate as responsible digital citizens are not acquired automatically and need to be learned, practiced 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 (WeProtect Global Alliance). Spunout argues that the issue of parental controls in online spaces cannot be separated from the need for effective media literacy measures and tools (Spunout).

A further description of media literacy: the ability to understand and critically evaluate broadcast, online and other media content and services – is, as Media Literacy Ireland argues, 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 (Brian O'Neill).

According to the Samaritans, user education and media literacy is a key facet of online safety, and they point to the UK Online Safety Bill whereby media literacy is underpinned by "an awareness of the impact material may have." This is a key principle of speaking safely about suicide and self-harm online. The 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. 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 (Samaritans).

The submission of the Rape Crisis Centres focuses on the importance of user empowerment. In their view the Code should: promote collaboration between the providers and educational institutions to promote digital literacy; and require providers to promote awareness among users of the avenues of complaint and redress available to them. 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 gender-based violence (Rape Crisis Centres).

SCU/ SLU state that 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 they recommend that the United Nations guidance for corporations on participation in child abuse prevention, guides this part of the Code (SCU/ SLU). This issue is further elaborated by DCU-ABC (see below).

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. They 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 the Commission to improve media literacy among audiences and content creators alike (Headline).

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. 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. Media literacy tools and resources available through VSPS must be age appropriate, easy to access and navigate (Bodywhys).

The submission from the OCO emphasises that media literacy tools, 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. 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 (Ombudsman for Children).

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 (Ombudsman for Children). Also emphasised by other stakeholders (DCEDIY, Safe Ireland, DCU- ABC, FSM, NICAM).

Conradh na Gaeilge states that media literacy measures and tools and awareness actions should not only be available in English but also available in Irish. It should be recognized that there is a community that uses Irish inside and outside the Gaeltacht. Therefore, if it is particularly important that there is public understanding of material published in print, broadcast, online or other media, it is necessary that the measures and effective media literacy tools are available in Irish as well as the English language (Conradh na Gaeilge).

The Children's Rights Alliance emphasises that this is an area that requires specific and targeted measures 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, and educational programmes and materials should be provided in order to develop digital literacy skills. 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 (Children's Rights Alliance).

Digital and media literacy programmes must be accessible to all groups and in particular the most vulnerable. According to research, Travellers experience low levels of literacy and low levels or media literacy as a result of exclusion within the education system resulting in low levels of school completion. Roma experience similarly low levels of media literacy and also face language barriers. 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.

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. 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. 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. 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. Hence, they recommend that: 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 (Children's Rights Alliance).

Several submissions also focus on media and digital literacy in relation to news and information. The Irish Safer Internet Centre believes media literacy is a crucial skill for all ages and given that 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 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. They also cite the 2023 Ofcom Children and Parents: Media Use and Attitudes report, which 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. 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 (Irish Safer Internet Centre).

Belong To emphasised that 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. They cite the Reuters Digital News Report, which found that online sources have remained the number one source of news information among the Irish public. 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%). 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". 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 (Belong To).

News brands 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. They are an active member of Media Literacy Ireland and 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 (News Brands).

The Dutch Regulator points to the important recommendations of the ERGA report from 2021: ERGA Media Literacy Report Recommendations for key principles, best practices, and a Media Literacy Toolbox for Video-sharing Platforms. Amongst other things, this report outlines how six key principles of media literacy initiatives can be implemented by VSPs. In addition, they also refer to the best practice in this field from the British regulator Ofcom (CvdM).

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 (FSM).

Google states that YouTube supports a flexible Code that emphasises effective media literacy and they believe that YouTube's best practice could be shared by the Commission 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 the platform, they seek to ensure users are not only consuming content but are also well-informed, critical thinkers. They are committed to these endeavours and continuously seek innovative ways to bolster media literacy, especially among young internet users.

A key focus of the media literacy measures has been the elevation of reliable and official information, thereby assisting users in differentiating between verified and potentially misleading content. These include fact-check feature that complement 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 YouTube 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 (Google).

Understanding the necessity for clear communication with users regarding safety measures and available tools, YouTube launched the "How YouTube Works" website in 2020. This platform offers lucid information on content policies, the repercussions of violating the 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 these platform-specific endeavours, YouTube is dedicated to fostering media literacy in broader contexts. They 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 collaboration with external experts (such as Barnardos in Ireland) and the independent assessments they conduct, ensuring they effectively alter the online behaviours of young individuals.

In November 2022, they also launched the '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, they 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 (Google).

Meta acknowledges 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.

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 dominion 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 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, Meta would encourage the Commission to make use of self- and co-regulatory solutions (Meta).

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. They consider this will be the most effective approach as different platforms will necessitate differing approaches depending on their scale, user base, content types etc. TikTok also provides detail of the activities they take in this area. They emphasise that they 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 users. They 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 navigate their online experiences safely and responsibly in connection with the platform. In order to raise awareness among users of specific topics and empower them, the run a variety of on and off-platform media literacy campaigns. The approach may differ depending on the topic. They localise certain campaigns (e.g., for elections) in that the collaborate with national partners and use language that the local audience can best connect with. They provide detail on both resources and specific campaigns.

Resources include: the Safety Centre providing 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; and a Help Centre providing accessible 'how to' explanations of the user experience to allow people to learn about the Platform and troubleshoot issues.

The "Safety" page contains the following subsections: the "Account and user safety" which further explains "Content violations and bans", the "Community Guidelines", Account Safety; and "user Safety"; and the "Report a problem" page which further explains how users (and non-users) can report content to TikTok (as well as reporting suspected underage users). Under the "Our Commitments section, there are user-friendly articles explaining TikTok's approach to Keeping People Safe, including separate articles that explains the approach to content moderation. They use Newsroom posts to

communicate with the community transparently and to build and maintain trust. They publish a range of posts in the Newsroom to generate awareness on safety and content related issues.

Specific campaigns have included: topics such as Covid-19, Covid-19 Vaccine, Holocaust Denial, MonkeyPox and the War in Ukraine. They use a combination of a number of in-app intervention tools such as video notice tags, search interventions and public service announcements. They have developed, together with fact-checking partners, and rolled-out eight localised (in Poland, Slovakia, Romania, Ukraine, Hungary, Estonia, Latvia and Lithuania) media literacy campaigns on the war. Users searching for keywords relating to the war are directed to tips, prepared in partnership with fact checking partners, to help users identify misinformation and prevent the spread of it on the platform. They 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 (TikTok).

DCU-ABC stated that 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). 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, a recent study in Norway, found that a large portion of children surveyed were not aware of Safety Centres. Hence, 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. It is recommended that the effectiveness of online safety and digital citizenship advice provided be periodically evaluated by the Commission (DCU-ABC).

DCU-ABC 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 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). The consider it 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-based and decided upon by experts in the field, rather than companies themselves. In a similar vein, they also believe that VSPS could be levied to provide financial support for organisations that administer helpline services. As previously documented 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 (DCU-ABC).

5.2 Terms and Conditions, Content Moderation and Complaints

This section provides an overview of the responses to questions 14, 15, and 16.

5.2.1 Terms and Conditions (Contents) – Measures (a) and (b)

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?

5.2.1.1 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?

Many responses focus on the need to emphasise illegal content and the consequences of posting, and respondents also list a range of harmful content that should be clearly defined and explained. Consequences should be clearly outlined. Industry responses reference the terms of the DSA that cover this issue.

Eurochild states that 'the code should pay attention to terms and conditions (T&C) when addressing contact risks and 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. The list of requirements should address all the harms related to children: content, contact, conduct and contract (including violence, discrimination, disinformation, sexual harm, commercial harm, manipulation, etc.). The Code could encourage the providers to enable children to adjust the content they want to have access (e.g., by blocking certain key words or tags)' (Eurochild).

WeProtect stresses that T&C should clearly state that a platform has a zero-tolerance approach regarding child sexual abuse content. T&C 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) (WeProtect, also IWF).

The DCU ABC stress that it would be important to go beyond the requirements of transposing AVMSD by regulating not only video but also text-based content. 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 (DCU ABC).

SCU/SLU proposes that content should be reviewed and categorised by moderators before being available to view. A safeguard is suggested where if a video is posted on a VSPS that it has to be filtered through a moderating system before being accessed by others. They also reference the recent Ofcom report "Regulating Video Sharing Platforms (CSPs): What we've learnt about VSPs' user policies" as valuable information on how to approach this and develop best practice models (SLU/SCU).

Women's Aid stress the importance of platform commitment to combat the spread of online violence against women and girls 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 violence against women and girls. This includes uploading or sharing of intimate images without consent. It should be clear that the consent of all people 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. It should be explained 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 recognise 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 explain steps that would be taken and commitment to short timeframes for action. The T&C should 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 (Women's Aid).

The Rape Crisis Centres emphasise a strict implementation of T&C, and state that 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. Providers should actively establish and utilise systems to identify repeat offenders of online abuse. Anonymity is a key tool utilised 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 (Rape Crisis Centres).

The Department of Health proposes to include prohibition of harmful content related to self-harm, suicide and eating disorders in the Code. 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. This is an opportunity for VSPS to explain why such content is harmful (Dept. of Health). Google notes that the education of creators is also key. More than 80% of creators who receive a warning never violate Google's policies again (Google).

Safe Ireland agrees with the list of prohibitions included in the Call for Inputs but believe 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. They think the default penalty for transgression of these prohibitions should be account suspension or termination. This is warranted in terms of the potential harm which may be caused (Safe Ireland).

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 have stated that they cannot publish 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. However, previously leaked moderation guidelines revealed serious omissions in how platforms regulated harmful online content. Therefore, the Code should either request from companies to reveal/publish operational moderator guidelines in the Community Guidelines, or the Commission should be able to request access to moderator guidelines from VSPS, at least for auditing purposes (DCU ABC).

Spunout emphasises that a minimum set of requirements for restricting online harms be mandated under the terms of the Code. They agree with 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 (Spunout).

The AVPA proposes the T&C 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 (AVPA).

In Google's 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 T&C and transparency of terms could be provided through non-binding guidance (Google).

Meta noted that Article 14 of the DSA already requires that all intermediary services have in place T&C 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 their 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. 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 T&C which comply with the DSA also comply with the Code (Meta).

TikTok references an EAO (European Audiovisual Observatory) Publication, which 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 TikTok's view that approach, as it is consistent with the majority of other EU Member States, is the correct approach. Adopting a different approach risks a patchwork implementation of the AVSMD across different EU jurisdictions and undermines the harmonised approach required by DSA (TikTok).

5.2.1.2 How should key aspects of T&C be brought to users' attention?

T&C should be presented in plain and easy to understand language (Department of Health, Children's Rights Alliance, Belong To, Rape Crisis Centre, Irish Safer Internet Centre), be concise, prominent and written in clear language suited to the age of children likely to access the service, (as per the 5Rights Foundation paper - Tick to Agree, Age appropriate presentation of published terms, September 2021). The 5Rights Foundation works with the Institute of Electrical and Electronics Engineers Standards Association (IEEE SA) to develop standards for Age-Appropriate Digital Services specifically aimed at promoting the rights and wellbeing of children in the digital world. Standard P2089 covers a broad set of issues: fair terms, children's rights, recognising children, platform liability and commercial drivers. Accordingly, organisations must consider the following:

Language: Published terms should avoid jargon and spell out key definitions and terms. 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.

When making a product or service inclusive and accessible, the following factors should be considered (IEEE Based on the 5Rights Principles for Children): 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; he 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." Obtaining meaningful consent means that a child understands and accepts the terms at all points and may choose to change their mind at a later point. 'Tick box' or 'unread' consent must not be used when the end user is a child. 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: T&C 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 (see stage 12 of IEEE Standard 2089 on age-appropriate T&C) (5Rights).

The Children's Rights Alliance and Belong To also reference the 5Rights paper and additionally propose that the Code should ensure: T&C should be accessible, transparent, fair, and available in child friendly language and/ or provide additional child friendly explanations in order to provide clear terms and policies; Terms of agreement should be proportionate to the value young people derive from the service; T&C 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; T&C must clarify what

happens when a user makes a complaint. They also reference the UK Children's Code (ICO) under the transparency standard, which states that if designated services need to draft their T&C 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 (Information Commissioner's Office, 'Age Appropriate Design: A Code of Practice for Online Services') (Children's Rights Alliance, Belong To).

The Irish Safer Internet Centre provides a response from the Webwise Youth Panel: "The T&C 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...". For members of the Webwise Youth Advisory Panel T&C is an important element and one that would benefit from proper consultation with young people and other vulnerable groups (Irish Safer Internet Centre).

The Rape Crisis Centre believes transparency and simplicity are key in bringing the T&C 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 T&C are brought to users' attention (Rape Crisis Centre).

T&C that are comprehensive but also transparent, easy to read and accessible to users are an essential feature of platform online safety. This approach is supported also by the BIK+ strategy which 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" (Brian O'Neill).

It is advised to include easy-to-read terms and relevant community rules or guidelines tailored to the needs of different age groups and supported by help articles and resources in a clearly identifiable Safety Centre or equivalent. In 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 (Brian O'Neill).

In the opinion of DCU-ABC, the Commission should request in codes that T&C 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 T&C /Community Guidelines/Standards content can include videos for younger children and not merely text (DCU ABC).

The HSE NOSP says 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 (HSE NOSP).

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; 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; 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; and the sharing of personal photographs of emaciation in order to seek approval and validation from peers (HSE NOSP).

The Samaritans align with the HSE NOSP position and add that in their experience content relating to self-harm and suicide often requires a more nuanced approach. 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, and to interact and build relationships that could help build their emotional resilience. According to the Samaritans 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 some types of content 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 – this can 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. The definition of 'online harms' should allow for these nuances in both the creation/uploading, and in sharing and consumption of the content. They 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 Guidelines by journalists and organisations in how they report on suicide leading to a reduction in sensationalised or dangerous news stories (Samaritans).

Women's Aid 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 (Women's Aid).

The ASAI supposes there is probably a tension between T&C 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 T&C, the VSPS can then rely on their agreement. The Code should require that an explainer of the key areas will be provided to users on screen and should not permit users to dismiss the content until a reasonable period of time has elapsed (ASAI).

5.2.1.3 *Examples of best practice in relation to T&C including content moderation policies and guidelines*

Spunout refers to what has been seen in recent months on the service formerly known as Twitter, the failure to effectively, consistently and permanently remove users who have broken the T&C 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 (Spunout).

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 repeated for the platform to take action) (DCU ABC).

Google YouTube's own policy framework shares common values with the safeguards of Article 28b of the AVMS Directive. Their 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 the platform must follow rules of conduct, including rules against sexualisation of minors, harmful or dangerous acts involving minors, inflicting emotional distress, and cyberbullying. They 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 Google Community Guidelines will result in a warning. Then they have a general three-strikes rule where three policy violations lead to account termination, but they 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 the services, they report it to the relevant authorities, and they disable the account (Google).

Meta believes that their 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. They have over eighteen years of experience in developing and enforcing content policies across their services and 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. They take a range of steps to help make the terms and policies of their services easy to understand and accessible and offer tools to help people make safe choices on their platforms and they work to be transparent about how they address these issues. Their policies – including the Facebook Terms of Service and Community Standards, the Instagram Terms of Use and Community Guidelines, and other specific policies (such as the Commerce Policies and Ad Policies) – are available online to everyone, via their Transparency Centre (Meta).

As a result, the 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 they enforce those policies and how they handle complaints from users in relation to their content enforcement decisions. 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 they enforce their policies. For example, when they enforce those policies against a user's account or content, they provide them with relevant information so that they can understand why Meta took action. Meta also releases a quarterly Community Standards

Enforcement Report (CSER), which shows how they are doing at enforcing their policies. This kind of transparency lets people see clearly how Meta is addressing safety issues and helps them get much-needed feedback (Meta).

TikTok considers that their Community Guidelines, their Terms of Services and other related documents (together, T&Cs) reflect best practice and would encourage the Commission to take these into account in the drafting of the Code. To assist the Commission in this regard, they have sought to outline below the elements of their T&Cs as representing aspects of the best practice in the industry and which they consider are relevant to any requirements under the Code. T&Cs and related documents should be easy to navigate and user-friendly. TikTok's 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. They use clear, simple and concise wording and make their T&Cs available in 25 European languages. They have also produced a summary of their T&Cs.

TikTok's Community Guidelines are organised by topic area, with each rule in bold. They first explain in brief what they don't allow, and they then provide more details, such as definitions and the range of actions they 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. TikTok's Terms of Service also includes a succinct and accessible "in short" section at the end of each provision, summarising the main points for users. They are always improving and evolving their policies and in their recent refresh of the Community Guidelines, they made several enhancements including more detail about how they use informational labels, warnings, and opt-in screens.

Online harms should be clearly addressed in T&Cs and related documents. TikTok's 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 the 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 the 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. Under the section Sensitive and Mature Themes (Article 28b(1)(a)) there is 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 (Tik Tok).

Permitted and prohibited use of the service (and consequences) should be clearly outlined in T&Cs. They believe TikTok's current approach of clearly but succinctly outlining in their 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 section of "What users cannot do on the platform", 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 the Community Guidelines and states that they "apply to everyone and to all content on the Platform". In their 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 they 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 (Tik Tok).

5.2.2 Applying T&C (Content moderation decisions) – Measures (a) and (b)

Question 15: How should we ask VSPS providers to address 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?

5.2.2.1 How should we ask VSPS providers to address content moderation in the Code?

The Children's Rights Alliance refers to the UN Committee on the Rights of the Child recommendation 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.' 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.' In their view, 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.' The Code should ensure that moderation is 'proportionate to the risk and activities associated with the product or service.' 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.' Moderation must be fair, unbiased and consistent for it to be effective. 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' (Children's Rights Alliance).

Regarding the accuracy of content moderation, Eurochild's view is that the Code should establish reporting requirements for video-sharing platforms. This would incentivise the companies to improve their content moderation practices iteratively and facilitate innovation. However, they know content moderation and self-reporting, especially among children, are not enough and lead to low levels of removal of harmful content. Therefore, they 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 (Eurochild).

The Irish Safer Internet stresses the importance of '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 but allow 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, they recommend establishing content moderation industry standards e.g., specialised training, must-have skill sets and expertise, staff-welfare and support, and quality assurance (Irish Safer Internet Centre).

In addition to previous proposals and positions, the SCU/SLU suggest that VSPS are supported through the provision of training modules to help their employees understand the impact of online harms on children. They believe that supervision of those who undertake the role of moderators of such content is the best way to support staff and ensure they maintain their knowledge and skills in implementing codes. Review mechanisms are needed to ensure that moderation is effective both in removing harmful content and reinstating content that was reported erroneously (SCU/SLU).

The matter of training moderators was also raised by Women's Aid. They align with the SCU/SLU's position on training modules, which should include the various forms of online violence against women. Moderators need to be culturally competent for the local areas they monitor. They need to also be trained in diversity and inclusion. For bigger platforms there could be specific Violence against Women and Girls 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. In addition, relevant violence against women and girls specialist services should be considered trusted flaggers in relation to image-based sexual abuse and other violence against women and girls 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 on violence against women and girls services could also have a role in informing the Commission about new trends in harmful violence against women and girls content (Women's Aid).

Safe Ireland's view is that it is challenging for moderators to make accurate decisions within as short a timeframe as possible. In order to mitigate this risk they suggest that this might be addressed by including a requirement in the Code that moderators should ask the person making the complaint 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 notes that borderline cases pose a challenge to moderators. They 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. They 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 (Safe Ireland).

When it comes to content moderation, The Dutch Ministry emphasises the point of deeper 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 (Dutch Ministry).

The German self-regulator FSM states that 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 (FSM).

Regarding content moderation, Google believes the Code should not be prescriptive about specific practices and instead focus on a principles-based approach as practices in relation to content moderation continue to evolve. 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. The 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 frameworks regarding terms and conditions disclosures, reporting mechanisms, user notice, internal appeals, out-of-court redress, and transparency. Google urge the Commission to ensure there is alignment and consistency between the Code and the DSA in this area (Google).

Meta states that the Commission should not prescribe requirements for content moderation. Content moderation is a complex, evolving and multifaceted approach that necessarily varies in detail across different services. 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 their view.

Regarding Article 15 of the eCommerce Directive, Meta agrees with the Commission. Article 15 indeed precludes the imposition of any general monitoring obligation on VSPS providers, and they note that Article 8 of the Digital Services Act also adopts the same approach. Meta respectfully notes 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 (Meta).

TikTok does not believe the Commission should address content moderation in the Code. They understand that one of the key objectives of the Code is the transposition of Article 28(b) of AVMSD and they 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, TikTok considers this to deviate from the provisions of the AVMSD. 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.

TikTok's position is 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 (TikTok).

The 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. 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 (ASAI).

The Department of Health stresses the importance of including harmful content related to suicide, self-harm and eating disorders in content moderation. They propose obligations which 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.

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 (Dept. of Health).

Belong To notes that effective content moderation ensures that the burden is not primarily placed on users to address harmful content through flagging mechanisms. It is important that the Code require social media platforms to follow best-practice guidelines in content detection and moderation, that are informed by LGBTQ+ identities and experiences. In recent years, media outlets have reported that VSPS have censored or suppressed LGBTQ+ content, creators and hashtags, despite this content not being in breach of community guidelines. 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 (Belong To).

News Brands and Local Media is concerned that their members who publish already regulated and trusted journalism, disseminated via an on-demand service, will become subject to policing by tech companies and their interpretations 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. News publisher content which touches on defined harmful online content categories risks being taken down or downrated 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 (News Brands and Local Media).

The Internet Watch Foundation proposes that 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 (IWF).

According to the Trust Alliance Group 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.

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.

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.

Signposting mental health support: Trust Alliance Group is 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 (Trust Alliance Group).

Headline reports a clear frustration among survey respondents about the quality and responsiveness of content moderation. 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 the Commission. If the Commission finds there is a track record of moderator assessment error, there must be actionable consequences for that VSPS. The Commission 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 (Headline).

The HSE NOSP opinion is that 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 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 (HSE NOSP).

The Samaritans raise the importance of the public being 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 the Commission on how to ensure that supportive content is not inadvertently removed (Samaritans).

The DCU-ABC suggest that for a high-level moderation the Commission should consult the Social Media Services Online Safety Code from the Online Safety Commissioner of Australia which applies to illegal content. 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 against

the policy; or otherwise sanction behaviours that contravene the T&C. In line with DSA requirements, the Code 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 Code should be in this regard (DCU ABC).

Spunout stresses that 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. Spunout would also state that they 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. Spunout feels 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 (Spunout).

Brian O'Neill's opinion on content moderation systems and processes is that they are the bedrock of the 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 a series of implementation reports on its Basic Online Safety Expectations – BOSE – process). The latter includes examples of reasonable steps a provider may be expected to take in dealing with a range of online risks (Brian O'Neill).

5.2.2.2 Are there any current practices which you consider to be best practice?

The German self-regulator FSM says that 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 (FSM).

YouTube uses a combination of automated and human evaluation to ensure content complies with their policies. In 2020, the most recent year for which they have figures, more than 20,000 people across the globe helped enforce Google's policies and moderate content. Their reviewer teams work around the world, 24 hours/7 days a week, speaking many different languages and are highly skilled. Their goal is to achieve both accuracy and scale in the work. Their flagging system allows their user community to notify them of any content that violates their guidelines and to help enforce policies. Moreover, they have also developed a "Priority Flagger" program to help encourage submissions of multiple high-quality flags about content that potentially violates their Community Guidelines.

As set out in the YouTube Transparency Report, between January 2023 and March 2023, they took down 8.7 million channels and 6.4 million videos on YouTube that failed to comply with their 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 they facilitate and encourage flags

by users, in practice, low actionability rates from user flags have required significant investments in Google's automated systems (Google).

Google 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 they balance taking action against content with respect for the rights to freedom of expression and access to information. Education of creators is also key. More than 80% of creators who receive a warning never violate Google's policies again (Google).

5.2.2.3 How should we address automated content detection and moderation in the Code?

The Children's Rights Alliance references the recommendation of the UN Committee on the Rights of the Child on automated systems 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.' 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' (Children's Rights Alliance).

The Women's Aid view on AI moderation is that it needs 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 timeframes (Women's Aid).

Safe Ireland agrees with the Commission's point on automated content detection that although it is 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. They 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 (Safe Ireland).

The Dutch Ministry believes that 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 (Dutch Ministry).

The German self-regulator FSM states that 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 fool proof. VSPS providers should be urged to have a robust human review process in place to ensure accurate and fair content moderation decisions (FSM).

Google recognises 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 (Google).

WeProtect believes 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. 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 (WeProtect).

According to the DCU-ABC, the 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 previous research into automated (Artificial Intelligence-based) moderation of cyberbullying, they 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. Recent research of the DCU-ABC involving 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. DCU ABC proposes to mandate transparency with respect to AI use and other automated means of risk/harm detection (DCU ABC).

In the context of effective content moderation, Spunout would encourage that the Code view automated content detection as a potentially useful tool for VSPS providers. 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 (Spunout).

5.2.3 Complaint Handling – Measures (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?

5.2.3.1 What requirements should the Code include about procedures for complaint-handling and resolution, including out-of-court redress or alternative-dispute resolution processes?

The 5Rights Foundation state that the requirements should include the provision of:

- 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;
- access to expert advice for children and parents access to support their decision- making and help them understand their rights.

Complaint-handling should also:

- 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 (5Rights).

They also note the recommendations of the UN Committee on the Rights of the Child and the Council of Europe Recommendation (Guidelines to Respect, Protect and Fulfil the Rights of the Child in the Digital Environment). These recommend 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. It further recommends that Member States should ensure that a child's right to an effective remedy under the European Convention of Human Rights is respected and protected when their rights have been infringed online. 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. Guidance is given on what constitutes an effective remedy and it includes elaboration on: inquiry; explanation; reply; correction; proceedings; immediate removal of unlawful content; apology; reinstatement; reconnection; compensation (mentioned by 5rights, Children's Rights Alliance, the Ombudsman for Children).

Belong To references the UN Committee on the Rights of the Child, which set out a number of recommendations relating to complaint handling and resolution. 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" (Belong To).

In addition, complaint and reporting mechanisms should be free of charge, safe, confidential, responsive, child-friendly and available in accessible formats. The 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 (Children's Rights Alliance).

The Irish Safer Internet shares the opinion of 5Rights regarding ease of reporting and clear routes for redress. In addition, they propose to include on option of reporting for non-account holders. They propose the Code should also include requirement on education about what to report, education about how reporting works, support reporting offline, listen to users on reporting and training for professionals working with children. They add insights from the Webwise Youth Panel. On the question 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, half of the students indicated they have never taken this action. When asked about the extra support people would need to step in and defend the targets of online bullying, 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 (Irish Safer Internet).

The OCO recommends that complaints-handling 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 which a VSPS provider responded to a report that the child made about alleged harmful content available on the VSPS. Informed by the Ombudsman's experience of dealing with complaints in the context of discharging their statutory complaints function, the OCO published a Guide to Child-Centred Complaints Handling in 2018. 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; seek feedback from children as part of a regular review of the complaints policy and procedures in place (Ombudsman).

Women's Aid expresses the opinion that the Code should require that platforms have clear complaint procedures, with appropriate timeframes, 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. When users report or complain about violence against women and girls content 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 violence against women and girls content, the appeal should be examined by a trusted service in the trusted flaggers scheme, or the Online Safety Commissioner (Women's Aid).

The Rape Crisis Centres raises the issue of intimate image abuse content (IIA). 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 Siochána 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. 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 (Rape Crisis Centres).

Safe Ireland suggests 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. 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 (Safe Ireland).

The general view of BFLGI is that 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. Regarding complaint handling and self-regulation, the BFLGI believes 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 (BFLGI and the Irish Heart Foundation).

The ASAI suggests that consideration be given to having high level requirements that a robust complaint 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. Where the subject matter of the complaint relates to commercial communications, 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 (ASAI).

The Trust Alliance Group 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. The Alliance believe this is an important step and are keeping this under review (Trust Alliance Group).

The opinion of Spunout is that 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, they 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. Spunout notes the statement that this consultation is not seeking views on whether the Commission should accept individual complaints, and welcome that this will be the source of a future consultation. They emphasise the strong belief that the Commission must adopt an individual complaints mechanism if successful implementation of the Code is to be assured. They 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 (Spunout).

5.2.3.2 To what extent should these requirements align with similar requirements in the DSA?

5Rights opinion is that regarding procedural aspects of complaint-handling and resolution, as well as reporting frequency of complaint handling systems and content requirements related to such reports, is that they should be aligned with requirements and procedures established under the DSA, so as to avoid duplications and ease compliance for VSPS providers (5Rights).

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). The German self-regulator FSM opinion is that complaint handling and resolution requirements should be consistent with those in the DSA to ensure consistency and harmonization across regulatory frameworks (FSM).

Regarding DSA alignment, YouTube has long allowed creators to appeal against their 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. Google recommends users go through their 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 (Google).

Google believes 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. Google would therefore welcome the Commission'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 (Google).

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. Google 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 the Commission should regularly audit these bodies' compliance against these standards (Google).

Meta reiterates the Commission statement 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 (Meta).

TikTok notes 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 how 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 (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, they 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 (TikTok).

WeProtect is of the opinion that the 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 and 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 (WeProtect Global Alliance).

5.2.3.3 What current practices could be regarded as best practice?

The Irish Safer Internet Centre recommends the practices as described in the following documents and Guides:

- 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 (Irish Safer Internet).

5.2.3.4 How frequently should VSPS providers be obliged to report to the Commission on their complaint handling systems and what should those reports contain?

5Rights view is that procedural aspects of complaint-handling and resolution, as well as reporting frequency such complaint handling systems and content requirements related to such reports, should be aligned with requirements and procedures established under the DSA, so as to avoid duplications and ease compliance for VSPS providers (5Rights).

The Children's Rights Alliance opinion is that the VSPS providers should be required to report on their complaint handling systems at a minimum annually (Children's Rights Alliance). The SCU/SLU recommends that VSPS should report to the Commission twice yearly and that the reports should detail complaints and responses as per the classification system they recommend above (SCU/SLU). Belong To is of the opinion that 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 the Commission (Belong to).

Women's Aid proposes that 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 timeframe in which they were solved. Complaints in relation to violence against women and girls content should be visible separately from other types (Women's Aid).

The Rape Crisis Centres also proposes that 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 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 (Rape Crisis Centres).

Regarding the frequency and the content of the reports to the Commission on complaint handling systems, Google launched their 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 their policies. This currently includes flagging (users and automated), video, channel, and comment removals, appeals and reinstatements, and highlighted policy verticals (Google).

Meta reiterates that 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, Meta has for many years published periodic transparency reports on their content moderation efforts and have worked to expand on this over time. They also publish data externally on a recurring basis on their response to government takedown requests and data requests. These efforts have been built upon to meet the DSA's additional transparency requirements (Meta).

TikTok outlines the DSA requirements for 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. Additionally, TikTok regularly publishes comprehensive voluntary Transparency Reports to provide visibility into how they uphold the Community Guidelines and respond to law enforcement requests for information, government requests for content removals, and intellectual property removal requests (TikTok).

5.2.3.5 *Should there be a maximum time-period for VSPS providers to handle user complaints and if so, what should that period be?*

The Children's Rights Alliance proposes that the process of handling user complaints should be speedy, child-friendly and provide the appropriate redress. 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 a quick and effective resolution for children and young people. They refer to the Online Safety Code developed by the Australian eSafety Commissioner, which states that Tier 1 social media services must resolve complaints within 'a reasonable time.' 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' (Children's Rights Alliance).

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 a complaint should be within 24 hours and should specify the next steps and how long they will take. The timeframe 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 (Women's Aid).

Similarly, the Rape Crisis Centres opinion is that in the case of harmful content of a sexual or intimate nature such as IIA 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 (Rape Crisis Centres).

Safe Ireland proposes 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 (Safe Ireland).

From the experience of German Self Regulator FSM under the German NetzDG they know that setting stringent timelines for complaint handling is challenging. While it is desirable from a user perspective that platforms review complains quickly, it is even more important that they make correct decisions in order not to limit 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 (FSM). The ASAI suggest that if maximum time periods are considered, it should be recognised that some complaints might relate to areas that should be prioritised, and flexibility around such time frames might be required (ASAI).

Regarding handling users complaints, Google emphasises that 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 DSA. Google 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 (Google).

Meta appreciates that the Commission wishes to hold VSPS accountable, but in their 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. Meta would encourage the Commission 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. Meta systems prioritise harmful content with the most views, which allows them to quickly remove content that is having the greatest effect on their users (e.g., terror content or content child abuse before looking at more harmless content types). Meta believes that the Commission 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. 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 the Commission chooses not to do so

(e.g., in the event of parallel complaints being raised under the complaint handling systems for each regime) (Meta).

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. The DSA imposes significant transparency obligations on intermediary services which includes information on complaints handling and as such TikTok considers 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 (TikTok).

VeryMy explains that it is a producer of the VerifyMyContent product where 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 (Verify my).

5.3 Possible Additional Measures and Other Matters

This section provides an overview of the responses to questions 17, 18, 19, 20, 21, 22 and 23.

5.3.1 Accessible Online Safety features

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?

Several respondents note that Article 47 of the Digital Services Act facilitates and encourages the drawing up of codes of conduct by the European Commission for the purposes of improving accessibility to people with disabilities. They state that it will be important that the Code take a non-prescriptive approach so as to ensure that the Commission's approach can be adapted in line with these codes of conduct once developed (Google, FSM, Meta, TikTok).

The Children's Right Alliance cite the UN Committee on the Rights of the Child concern that children with disabilities may be 'more exposed to risks, including cyberaggression and sexual exploitation and abuse, in the digital environment.' The Committee has recommended 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. This should be done in a way that counters 'prejudice faced by children with disabilities that might lead to overprotection or exclusion.' It is important that information is provided in accessible formats on safety and protective strategies (Children's Rights Alliance, Irish Safer Internet Centre).

One method of ensuring this is using equality proofing safety measures and providing guidance on various accessibility methods in place. The Children's Rights Alliance emphasises that the Code must respect the evolving capacities of all children including those of children with disabilities or in vulnerable situations. 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. Hence they recommend that: information should be provided in accessible formats on safety and protection strategies; safety measures should be equality proofed as a matter of standard practice; and 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 (Children's Rights Alliance, Irish safer Internet Centre, SLU /SCU).

WeProtect provides results of their 2021 research, conducted by Economist Impact, titled "Estimates of childhood exposure to online sexual harms and their risk factors." The research 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. In an additional Briefing Paper on the sexual exploitation and abuse of deaf and disabled children online, WeProtect stated 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. VSPS 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. VSPS 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 (WeProtect Global Alliance).

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 environment" 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'.[...] Also, "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 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' (Irish Safer Internet Centre).

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 (WeProtect Global Alliance). 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 (Brian O' Neill).

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 (Irish Safer Internet Centre). Several other respondents also stressed that accessibility of the safety functions represent a key element of safety by design (Brian O' Neill, National Parents Council, Women's Aid).

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" (with reference to the Disability Act 2005) (Brian O' Neill).

According to the National parents council, 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 (NPC).

Women's Aid has pointed out that flagging systems should also be designed with the needs of children, young people and people with additional needs and/or disabilities in mind. 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. 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) (Women's Aid). 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, 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 (Rape Crisis Centres).

With respect to psychosocial disabilities, it is important that the Code empowers those who have identified triggers individual to their mental health condition and 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, after 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 (Headline).

5 Rights addressed accessibility in relation to the development of terms and conditions, which need to be accessible and 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: 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 (5Rights, also cited by Belong To).

Several respondents make reference to the UN Committee on the Rights of the Child 2021 General Comment which recommended 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'. The Committee also recommended that 'complaint and reporting mechanisms should be free of charge, safe, confidential, responsive, child-friendly and available in accessible formats' (Belong To, Ombudsman for Children).

The Ombudsman mentions a guide that they published in 2018, the Guide to Child-Centred Complaints Handling. 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 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 (OCO).

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 (FSM). The Internet Watch Foundation also agrees that terms and conditions, user safety functionality should be easily comprehendible to all users. Best practice in this area could include easy read versions of terms and conditions (IWF).

Google outlined the measures that YouTube has in place to ensure that the service is accessible to people with disabilities. They hope that this information will assist the Commission 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. Google explained that they try 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 they develop and innovate their products; however, they 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.

YouTube continues to innovate to deliver new features. For example, they have recently redesigned icons on the app to be readable, consistent and clearly understood. Thanks to long term investments in machine learning, they now provide automatic captions in more languages. They 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. The app also works with Android features and informs users how to turn on or disable features that can aid with app usage (Google).

5.3.2 Risk Assessments

5.3.3 Safety by Design

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?

5.3.3.1 What approach do you think the Code should take to risk assessments and safety by design?

Regarding child safety, 5Rights proposes that VSPS providers should assess the risks outlined in the 4Cs framework (Content, Contact, Conduct, Contract) presented by each feature of the product or service to reveal known harms, potential risks and unintended consequences. 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. 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. 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.
- Misuse of features: VSPS providers should account for how their features might be misused by actors with malign intent.
- Risks over time: Certain risks may expose children to low levels of immediate harm but increase in severity over time.
- Children's vulnerability to harm: VSPS providers should consider the particular vulnerabilities of different children, at different development stages.
- 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 (5Rights).

The Children's Rights Alliance refers to the COE's 'Recommendation Guidelines to Respect, Protect and Fulfil the Rights of the Child in the Digital Environment'. 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. Including these principles in the Code would help ensure that, from the planning stages of technology development onward, children are protected. They cite the 5Rights research findings 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.' Therefore, the Code presents a huge opportunity to embed the principle of safety by design into the Irish regulatory framework. The Children's Right Alliance proposes that: (i) the Code should require platform operators to regularly undertake child-rights impact assessments in relation to digital technologies and demonstrate that they

are taking reasonable steps to mitigate risks (this in line with the Council of Europe Guidelines), (ii) child rights risk assessments should be conducted before their digital products or services could reach or affect children and (iii) businesses should be obliged to "undertake child rights due diligence, which entails that businesses should identify, prevent, and mitigate their impact on children's rights including across their business relationships and within global operations" (Children's Rights Alliance, also referenced by Belong to).

The opinion of Cybersafe Kids on requirements for 'safety by design' in the Code is that 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. They are not convinced that VSPS providers will truly adopt all necessary measures for a safety by design approach unless compelled to do so (Cybersafe Kids).

Eurochild emphasises that 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 as that of the DSA but applied to the content in scope of the Code, integrated as much as possible with the requirements of the DSA that the provider might have to comply with. For VSPS who are not VLOPs, a reduced version of such an assessment could be required. These risk assessments should include or be complemented with a children's rights impact assessment. Regarding safety-by-design, they strongly recommend that the Code should focuses on some minimum requirements on safety and privacy by design. 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 (Eurochild).

BFLGI and the Irish Heart Foundation say that child rights impact assessment (CRIA) should be mandated in the Code. 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) (BFLGI and the Irish Heart Foundation).

The Department of Children, Equality, Disability, Integration and Youth (DCEDIY) supports the proposal to require VSPS providers to follow a 'safety by design' approach when they introduce new features. They 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. DCEDIY supports the proposed requirement to prepare a "Safety Impact Assessment" whenever services are being developed or enhanced, with a sign-off of the risk assessment and proposed mitigation measures by an executive staff member of the VSPS provider with appropriate experience and responsibilities. The DCEDIY position is also supported by the obligations contained in the Children First Act 2015 and the Addendum to Children First Act: National Guidance for the protection and welfare of Children (Keeping children and young people safe from harm online) (DCEDIY).

The SCU SLU position is that '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 (SCU SLU).

Women's Aid proposes a list of principles to be taken into account in design:

- 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 recognises 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 image-based sexual abuse content to relevant supports in their country.
- Highlight no tolerance of violence against women and girls content and image-based sexual abuse content 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 image-based sexual abuse 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 violence against women and girls online and image-based sexual abuse 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 (Women's Aid). The Rape Crisis Centres state that 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 (Rape Crisis Centres).

Samaritans Ireland believes that the prevalence and placement of harmful online content, particularly suicide and self- harm 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. Regarding 'safety by design' they emphasise that algorithms to deliver content on the basis of use / clicks / reach 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 (Samaritans).

WeProtect draws attention to the fact that 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: (i) The dissemination of illegal content, (ii) Negative effects for the exercise of fundamental rights, (iii) Negative effects on civic discourse and electoral processes, and public security, and (iv) 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 (WeProtect).

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". They understand that the professional body representing internet service providers in Ireland, Hotline.ie, itself favours this approach (Safe Ireland).

The German self-regulator FSM is of the opinion that 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 (FSM).

NICAM's position on safety by design is that it is important to be transparent about the platforms' algorithms. For children, highly personalized algorithms should not be used at all as they pose too many risks for this vulnerable group (NICAM).

The industry stakeholders (Google, Meta and TikTok) are unanimous regarding the approach to risk assessment in the Code. They say that 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 (Google).

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 systemic risks stemming from their platform and services:

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.

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) (TikTok).

VLOPS will also be required to provide reports on the risk assessments to relevant supervisory authorities, in Meta's case, to the EC and the Commission (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). 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. 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 services (TikTok). As such, the industry urges the Commission to ensure that the Code is aligned to avoid multiple requirements for risk assessments which cut across the harmonised approach of the DSA (Google, Meta, TikTok).

IWF is supportive of the provisions in the Digital Services Act that focus on Very Large Online Platforms (VLOPS), but claim 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 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 threats and risks 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) (IWF).

The DCU ABC restates that the requirement for companies to conduct risk assessments is a provision for VLOPs in the DSA (Article 34) and they 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 their 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) (DCU ABC).

AVPA highlights that privacy-by-design and data minimisation are already required under GDPR. The Commission should emphasise this requirement for age assurance solutions to reassure the public about data security and privacy. AVPA's Code of Conduct seeks to export European best practices to their members globally and is a requirement of full membership of their Association (AVPA).

UCD proposes to conduct risk assessments through audits of algorithms and data on VSPS. Access to platforms must be mandated in order that third-party auditors can conduct these audits (UCD).

5.3.3.2 Are there any examples you can point us towards which you consider to be best practice?

The Irish Safer Internet Centre references the model of 'Child Rights by Design' proposed by the Digital Futures Commission and the 5Rights team. The model provides a toolkit for designers and developers of digital products and was co-developed with them – and with children. 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 world.

The other example of good practice is the Australian eSafety Commissioner's Safety by Design 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 rewarding online experiences for everyone (Irish Safer Internet).

In addition, the Australian eSafety Commissioner says the 'Safety by Design' is a voluntary, non-regulatory measure in Australia. 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 the industry to imbed Safety by Design within product development and deployment. 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 (Australian eSafety).

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 (SCU/SLU).

The Netherlands is developing various tools to better protect children online. These instruments inform 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) that 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. 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. The Netherlands 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 (NL Ministry).

DCU ABC recommends consulting the Australian Online Safety Commissioner's Basic Online Safety Expectations (BOSE) 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. They 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 (DCU ABC).

Brian O'Neill also recommends the Australian eSafety Commissioner's Safety by Design Overview. 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 assessment 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 in the context of digital products and services for mobile devices. The Digital Futures Commission (and 5Rights Foundation) has developed a dedicated resource called Child Rights by Design containing guidance for innovators when designing digital products and services. Separately, 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 (Brian O'Neill).

5.3.4 Cooperation with other Regulators, Bodies

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

As regards cooperation with other regulators and bodies to help with the implementation of the Code, the responses focused on both international and national regulatory cooperation, and also on cooperation at the national level with a range of statutory bodies and other organisations and NGOs.

The Dutch regulator emphasised that 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 VSPSs. 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. ERGA tries 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. 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. They are currently indexing these practical challenges within the 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 2023. It would also be beneficial to establish a European VSPS 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 VSPS Regulation group within EPRA, which provides workshops, is a good example of such a group (CvdM).

Ofcom emphasises that regulatory coordination and cooperation benefits regulators and helps them to further their respective domestic objectives. It is also helpful for the services being regulated, and coordination around regulatory expectations and supervisory approaches can promote services' compliance across jurisdictions. Ofcom notes that a range of VSPS that are used by UK-based adults and children are beyond their regulatory remit and fall under the remit of counterpart regulators in other European countries. As such, to protect individuals in the UK, Ofcom depends on regulators like the Commission, just as they rely on Ofcom to protect individuals in their countries from safety

challenges that arise on the VSPs that are notified to Ofcom. In this context it is essential for counterpart VSP regulators to build and maintain close working relationships to (among others): work towards consensus on what regulators consider to be the appropriate use of safety measures; collaborate to create a common understanding of the criteria and standards by which regulators evaluate the effectiveness of these measures; improve safety outcomes and reduce compliance burdens for providers across jurisdictions; work towards coordination of investigations and enforcement; and hence 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 VSPS regulators (Ofcom).

Cooperation among international online safety regulators and the multistakeholder policy community is a key ingredient for the success of this new era of regulation. There is not yet an international consensus on how these new concepts such as risk management, codes of practice, and mandatory transparency reporting should be implemented. Regulators can work together towards common understandings of the norms, principles, and standards that will determine how these regulatory tools evolve in the years to come. The Ofcom regulatory philosophy is to aim for international alignment with regulatory partners where appropriate and possible. As in the specific example of VSPS regulation, there are many benefits to this approach – for Ofcom, other regulators, the companies regulated, and ultimately the users that the online safety regulation seeks to protect. They welcome the fact that the Commission 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 they can together shape a proportionate, effective, and rights-respecting global approach to online safety regulation (Ofcom).

The German self-regulatory body states 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. They 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 (FSM).

According to NICAM, they believe that ERGA could play a good role in the coordination of an international approach to content rating and information on VSPS. Their view on this is that an international system for rating productions on VSPS should be implemented. In this system each member state 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 (NICAM).

Google stressed that 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 (Google).

Meta reiterated that they 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. They also believe that it would be beneficial for the Commission to cooperate with the UK's Ofcom which also oversees a comparable (though not identical) VSP regime (Meta).

According to Technology Ireland, the reputational risk for Ireland cannot be underestimated with the successful rollout of the Code. The Commission 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 the Commission should position itself to offer stable guidance in this regard (Technology Ireland).

The Irish Safer Internet Centre notes that 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 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 (Irish Safer Internet Centre).

It is noted that cooperation takes on particular significance within the EU, 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 Network has similarly been valuable for knowledge exchange, particularly given the early stage of development of regulatory practice in this area (Brian O'Neill).

It is also emphasised that 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 (WeProtect Global Alliance).

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 (Women's aid).

According to Safe Ireland, 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 future cooperation with the other EU Digital Service Coordinators and the European Commission 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 is important for the Commission to maximise its opportunities in this regard (Safe Ireland).

The IWF agree that collaboration with other regulatory bodies, such as those outlined in the consultation will be important and they encourage the involvement of relationships with the global regulators network and ERGA. They also recommend that the Commission develop relationships with service providers such as the IWF whose datasets 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 (IWF).

The Samaritans noted the significance of cooperation following the death of Molly Russell in 2019. The 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 document 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 (Samaritans).

The Australian eSafety Commissioner cites the Australian example of cooperation. In March 2022, recognising the importance of collaboration, 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). 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 various respective regulatory agendas - and the ability to identify and anticipate opportunities for coordination and tensions that may arise. eSafety recognises that they are part of an international regulatory ecosystem, and welcome new regulators and novel approaches being tested to improve online safety globally. eSafety is working with international partners to share what they learn – and to learn from their planned approaches, through the Global Online Safety Regulators Network and bilaterally. The eSafety Commissioner's office is delighted to be collaborating with the Commission, because they 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 regulators are able to tackle emergent threats (Australian eSafety Commissioner).

Meta stressed that to the extent that the Code touches on issues which may overlap with 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 the Commission 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, they believe that the Commission should involve these regulatory bodies in developing relevant codes in consultation with industry (Meta).

According to Google, 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 the Commission to align its approach with other regulators, for example the Data Protection Commission (Google).

TikTok noted that 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) (TikTok).

The Department of Health 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 (Department of Health). According to the HSE National Office for Suicide Prevention, 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 Code and mechanisms arising across different VSPS providers (HSE NOSP).

The Children's Rights alliance emphasised the 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 (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 National Disability Authority should be coordinated with. In conclusion, they recommend cooperation with other public bodies and government departments including Tusla, IHREC, DCEDIY, and the National Disability Authority in order to ensure effective implementation of the new Online Safety Code (Children's Rights alliance).

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 (Irish safer internet).

The Irish Heart Foundation and the BFLG share the opinion that 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. 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 (Irish Heart Foundation, BFLG).

Several organisations emphasise the importance of on-going dialogue with all stakeholders. 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 (SCU/ SLU). 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 (RCNI).

The response of Cybersafe Kids discuses Bloggers and influencers on VSPS who generate user content and frequently use their children to promote the page, the brands or activities. 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. They emphasise that there should be cooperation between the OSC and the CCPC in requiring that these brands act responsibly when entering into such agreements with parents as, for example, the child's privacy is often compromised. They emphasise the need for cooperation regarding obligations on fair division of earnings, rest breaks, prioritising the child's welfare and ensuring that the child's privacy can be maintained (Cybersafe Kids).

5.3.5 Harmful feeds and recommender systems

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?

5.3.5.1 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?

The majority of non-industry stakeholders agree on a general concern that recommender algorithms 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 (National Parents Council, Irish Safer Internet Centre).

5Rights opinion is that for minors the recommender systems should be off by default. Certain providers of online services are already 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. The 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 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.
- 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 (5Rights).

Eurochild’s position is that 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 (Eurochild).

The Irish Safer Internet Centre refers to the office of the eSafety Commissioner Australia position – that 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 illness. In addition, risks can be greater for children and young people, especially if they receive:

- friend/follower suggestions encouraging them to interact with potentially dangerous adults
- content that encourages binge consumption without breaks
- content that promotes ‘ideals’ of body types and beauty stereotypes
- 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.

The Irish Safer Internet mentions the Pathways report (How digital design puts children at risk), which is the outcome of a research project undertaken by Revealing Reality on behalf of the 5Rights Foundation. It examines how design choices embedded in digital products impact the lives of children. It should also be noted that recommender systems and algorithms have many benefits to users for example, with more targeted search results, discovering new information, and providing a more personalised experience online (Irish Safer Internet Centre).

In addition to a general concern that recommender algorithms can pose several potential risks for children and young people, the SLU/SCU focuses on the areas of child sexual abuse and mental health. It is their view that mechanisms to address this need to be developed. They 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. A mechanism whereby parents would be sent summaries of websites accessed, such as is currently used by service providers to inform users of average screen time in a day or week could be used (SLU SCU).

Women's Aid proposes the inclusion of cross platform cooperation. They say platforms should 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 (Women's Aid). The Rape Crisis Centres opinion is that it is vital for providers (and their design staff and moderators) to have an evidentially informed understanding of GBV (Gender Based Violence) to be able to design safety features and assess user complaints effectively (Rape Crisis Centres).

The RCNI (Rape Crisis Network Ireland) submits 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. They 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 (RCNI).

The Dutch Ministry draws attention to the fact that 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. The NL ministry is hoping and expecting that these regulations will help protect (minor) users against the harm caused by the aggregate impact of content. However, they would be interested in seeing if Ireland is planning to implement any additional requirements regarding this topic (NL Ministry).

The Irish Council for Civil Liberties (ICCL) emphasises the following facts:

- 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". In addition, platforms are required by Article 38 DSA to provide alternative recommendations not based on a profile of the user.

- 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

The Irish Council for Civil Liberties opinion is that 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. They point to 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 (ICCL).

The recommendations of the ICCL are outlined below:

- 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.
- 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 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.
- 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 (ICCL).

The Department of Health 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 (Dept. of Health).

WeProtect proposes interventions by the platforms, which could include prohibiting harmful 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 (WeProtect).

The HSE NOSP raises the issue that 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-harm. Several respondents to a UK hospital-based qualitative study admitted to intentionally seeking information about methods when planning their attempt — predominantly from the internet. 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 (HSE NOSP).

The Samaritans raise the issue of smaller platforms where some of the most harmful suicide and self-harm content exists. 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. Therefore, 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 (Samaritans).

The UCD response proposes 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 (Convolutional Neural Network, deep learning) can be used to generate representations for images and videos. Such analysis can be conducted by third party auditors, but this relies on them having access to algorithms and data or access though an API (UCD).

Brian O'Neill's opinion is 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 (Brian O'Neill).

5.3.5.2 Are there current practices which you consider to be best practice in this regard?

The Irish Safer Internet states that 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. They consider some current practices 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 (Irish Safer Internet)

The Dutch regulator, the Commissariaat voor de Media (CvdM) shares information on bad practices with regard to recommender systems. The non-profit Centre for Countering Digital Hate (CCDH) published a paper, 'Deadly by Design', 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' (CvdM).

The Irish Traveller Movement agrees the Code should require VSPS to ensure their recommender systems do not result in a feed of content which in aggregate risks causing harm as is the case in examples of 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. They recommend establishing an 'at-risk' advisory group to work with and inform the Commission's undertaking of the model and to include content providers (Irish Traveller Movement).

Meta Platform Ireland Limited is confident that it has pioneered many best practices for the benefit of users (and in particular, younger users). For example, they refer the Commission to MPIL's youth safety strategy. 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. They have launched 'System Cards' - 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 their 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-personalised experience. Meta considers that this will further address any potential negative effects that the Commission considers might arise from the aggregate impact of content. In addition, they 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 (Meta).

TikTok works continuously to improve their recommender systems to, not only improve their product, but also to develop new strategies to interrupt repetitive patterns that may include harmful content.

- Their 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, their 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 the 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. TikTok also test ways to recognise if their 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 TikTok's Content Levels system.
- TikTok has introduced a 'refresh' feature that enables people to refresh the For You feed. When enabled, this feature allows someone to view content on their For You feed as if they just signed up for TikTok. TikTok's recommendation system will then begin to surface more content based on new interactions. This feature adds to a number of content controls their community already has to shape their experience. For example, people can choose to automatically filter out videos that use specific hashtags or phrases from 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. Enabling refresh will not override any settings a user has already chosen to enable or impact accounts they have followed.
- As a result of TikTok's continued testing and efforts, they constantly improve their platform viewing experience, so viewers now see fewer videos about these topics at a time. They are also working to recognise if their system may inadvertently be recommending only very limited types of content that, though not violating policies, could have a negative effect if that is the majority of what someone watches, such as content about loneliness or weight loss. TikTok's goal is for each person's For You feed is to feature a breadth of content, creators, and topics. This work is being informed by ongoing conversations with experts across medicine, clinical psychology, and AI ethics, members of TikTok's Content Advisory Council, and their community (TikTok).

The industry responses on how to address the problem of feeds which cause harm because of recommender systems is as follows: Google expects that 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 (Google).

Meta's position is that 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, the Commission should take into account that the EU legislature chose to exempt non-VLOPs from those obligations (Meta).

TikTok recommends that the Commission should adopt a principles-based approach (supplemented by guidance if necessary) thus allowing for platforms like TikTok to iterate their methodologies and technological solutions. 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 (TikTok).

5.3.6 Audiovisual commercial communications arranged by the VSPS provider

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?

There was a broad discussion of harmful commercial communications under Chapter 3 above, and further discussion under chapter 5 regarding commercial communications in content uploaded by users. In addition, the issue of influencers was also discussed above. This section is focused on the obligations for VSPSs regarding the commercial content arranged by the VSPSs themselves. Responses here referred to obligations under the Audiovisual Media Services Directive and the Digital Services Act. In addition, the relevant provisions of the OSMR Act were cited in relation to standards, and international guidelines were also referenced.

With regard to the commercial content offered by VSPs, the Dutch regulator stressed that this 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. 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. Where 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 (Cvdm).

Furthermore, all these measures must also take into account Article 28b (3) of the AVMSD that minors should be protected against inappropriate advertisements without 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 they 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 (Cvdm).

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 futureproofing of standards given the fast pace of change in technology platforms and commercial practices (Brian O' Neill).

As per the guidelines of the AVMS Directive, "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". Google believes 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 (Google).

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, META considers 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 promo□ng 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. Its 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. How those policies are enforced may also be taken into consideration by the Commission in the Code. For instance, their 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. Meta uses automated and, in some instances, manual review to enforce its policies and, beyond reviewing individual ads, Meta also monitors and investigates advertiser behaviour, and may restrict advertiser accounts that don't follow the advertising policies, Community Standards or other 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) (Meta).

TikTok states that in line with their general comments to the consultation, they 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. 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. 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. The Community Guidelines and Ad Policies are enforced using a combination of automated and human moderation (TikTok).

According to Alcohol Action Ireland, it is essential that any new codes which are developed by the Commission must ensure that children are not targeted by alcohol advertisers either in online or traditional broadcast marketing. 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, 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 (Alcohol Action Ireland).

The Children's Rights Alliance recommend that consideration should be given to addressing Harmful Commercial Communications, particularly marketing of high fat, sugar and salt foods, breastmilk substitutes and alcohol. 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. In addition, 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 (Children's Rights Alliance).

SLU/ SCU state that VSPS providers should adhere to international best practice, such as the UN Guiding Principles on Business and Human Rights and this requirement should be reflected in the Code. Transparency regarding commercial content is required (SLU/ SCU).

The Advertising Standards Authority of Ireland placed an emphasis on industry self-regulation. As the VSPSs 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;
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;
d) and 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. They state that their 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. They 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. They also note the AVMSD Article 9 (4) in relation to HFSS foods that encourages the use of co- and self- regulation. It was 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 4a(1) (ASAI).

5.3.7 Compliance

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

It is important to hold platforms accountable through effective evaluation and monitoring of complaints and reports, made publicly available. There should be built in positive reinforcement mechanisms for VSPS with good compliance (Headline).

Google notes that the AVMS Directive requires the Commission to assess the appropriateness of the measures that VSPS providers take under Article 28(b). They would welcome a monitoring framework that is proportionate and focused on structural compliance rather than prescriptive reporting

requirements. They 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 the relevant outcomes to be achieved, with different VSPS being given scope to determine how best those outcomes can be achieved. They 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 (Google).

META reiterated comments from its 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, they would ask that the Commission 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:

- Per Articles 15, 24 and 42 of the DSA, extensive periodic transparency reporting is required (see response to Questions 9 and 16).
- 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 Queston 18).
- 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 the Commission, as Meta's Digital Services Coordinator. Such assessments and reporting requirements under the DSA are significant and extensive and 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 the Commission 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. They respectfully request that before determining any reporting obligations under the Code, that the Commission consider the possibility that the data which may be needed to perform their 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 (Meta).

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. As regards the compliance monitoring and reporting arrangements that the Commission should include within the Code, TikTok notes 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, they 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. They 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. 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) they 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 (TikTok).

Verifymy details products that they have developed to assist with gathering data to provide reporting. With the power of artificial intelligence, they 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 (Verifymy).

The IWF state they would be happy to consider assisting with compliance monitoring and reporting arrangements, by providing data that they have on the extent of harm on platforms, and their annual report, which already provides detailed information on most of this and could have a role in helping to evidence the prevalence of child sexual abuse online. They would not, however want 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 to IWF in further understanding the prevalence and how effective these services are at preventing viewing offences or the upload and further distribution of this illegal imagery (IWF).

Other responses discussed reporting in relation to best practice. 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. 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 sharing services have been published in respect of TVEC53 and shortly in relation to online child sexual abuse and exploitation (forthcoming). 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 (Brian O’ Neill).

According to We protect, 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 (We protect).

Spunout strongly recommend that compliance with the 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, they believe that full compliance with the Code should be a matter for annual review and sign-off by VSPS Boards of Directors. This should 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 (Spunout, Irish Safer Internet Centre).

The Irish Safer Internet Centre recommends that the Code to refer to "The Fundamentals" document of 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). Hence, as deemed appropriate within the purpose, scope, and audience, it is recommended to include the following information, non-exhaustive and in no particular order:

- the nature, context and content of the relevant material and the severity of its impact and harm;
- the extent of avenues available or suitable to address the type of harm and whether such approaches have been successful or not;
- harm reduction indicators and measurement;
- preventive and deterrent measures deployed and the effectiveness in harm reduction;
- whether the intended subject of the regulatory action has been the subject of prior compliance or enforcement action, and the outcome of that action;
- the extent to which any conduct represents a broader systemic issue;
- the circumstances of the end-user any indicators of vulnerability and level of support required to respond to compliance or enforcement action;
- emerging trends and issues identified during the reporting period;
- the action times on complaints handling broken down per type of resolution. (Irish Safer Internet Centre)

Women's Aid outline the type of information needed on reporting in relation to violence against women and girls content and image-based sexual abuse content, VSPS should be required to monitor and report quarterly to the Commission on:

- Preventative measures taken to limit violence against women and girls online and in particular to prevent the spreading of image-based sexual abuse content, including risk assessment carried out.
- How many trained moderators they have available to monitor these issues specifically.
- Number of image-based sexual abuse / misogynistic videos flagged, outcomes and timeframes.
- Number of complaints received, the outcomes and timeframes.
- Number of videos promoting violence against women and girls removed.

- Number of videos with image-based sexual abuse 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 (Women's Aid).

DCU-ABC state 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). 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, 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 (DCU-ABC).

According to the Irish Traveller Movement, 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'. 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', it recommended that there by a 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) (Irish Traveller Movement).

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 (BFLGI and Irish Heart Foundation).

On the issue of monitoring and enforcement, the BFLGI and the Irish Heart Foundation 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 the online safety Code.
- Use of existing protocol Netcode to ensure careful creation of the online safety Code.
- Regulated entities should not just be required to "provide periodic reports on their compliance or otherwise with codes" but should also be obliged 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 (BFLGI and Irish Heart Foundation).

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 (RCNI).

The DCU-ABC response reiterated their belief that 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 they 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. While such provisions already exist in the OSMR as regards auditing company activity, they 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 (DCU-ABC). According to SLU/ SCU external audits of VSPS are required to ensure monitoring and that reporting arrangements are adhered to (SLU/ SCU).

Safe Ireland emphasised how important it is 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 adult victims of domestic abuse) when it comes to compliance, monitoring and reporting

arrangements. They 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, they 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 (Safe Ireland).

According to the submission of the Australian eSafety Commissioner, the law allows them to also require reports from the online services. The Commissioner can issue a notice for a report on specific issues or a specific platform. This must be based on a set of criteria: including 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. For example, 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 also has powers to require periodic reporting to track key issues and metrics over time (Australian eSafety Commissioner).

The Samaritans 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 well-being 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 key to monitoring internet safety. The monitoring/report should include specific measures for platforms to ensure the good mental health/wellbeing of people who review/moderate potentially harmful content. It should include 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 (Samaritans).

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'). 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. 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 to adopt an online safety code that is effective, practicable and acceptable to all (Rape Crisis Centres).

5.3.8 Transitional Arrangements

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?

Regarding transition periods, most respondents recognised that some level of transition was necessary. However, many urged that this be as short as possible, in particular as they claim that the major players already have some systems in place, particularly due to the DSA. In addition, they state that the VSPS have been aware of the upcoming Code and the likely content of the Code. The responses include various suggestions for an appropriate transition period. It is also emphasised that the time needed for implementation will depend on the detail of the Code.

The Dutch regulator states that 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 (CvdM). Likewise, NICAM states that 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 (NICAM).

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 (Brian O' Neill).

It is the opinion of the SCU/ SLU 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. They 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 (SCU/ SLU).

A similar response was provided by the RCNI, who stated that 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 a learning period in the roll out of a Code no matter how well crafted. They 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 (RCNI).

The Department of Health states that 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 (Department of Health).

According to the WeProtect Global Alliance, 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 (WeProtect Global Alliance).

Spunout questions 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, they 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 (Spunout).

The Children's Rights Alliance stressed that 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. They believe 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. 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. A timeframe like this could be considered (Children's Rights Alliance).

According to the response of Belong To, 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. Research conducted in 2021 found that LGBTQ+ people experience 50% more online hate and harassment than any other minority group. 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 (Belong To).

Internet Watch Foundation note that 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. They 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. They note that with the

development of the Digital Services Act, the enforcement aspects of the regulation have taken around 12 months to come into force (Internet Watch Foundation).

According to the Age Verification Providers Association, 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 AVPA 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 (AVPA).

Google 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. They would urge the Commission 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 the Commission 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, they 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. They look forward to working with the Commission as it develops the Code to determine whether, and to what extent, such a period is required (Google).

TikTok agree with the suggestion that the Code should have an appropriate transition period. Given the issues addressed by the Code are necessarily intertwined and their view that the Code should set out high level principles, they 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, they are not in a position to provide more specific guidance or assistance to the Commission. However, once the Code is close to completion, they would recommend the Commission engage with industry on this issue. In light of the fact that providers will have undergone and are still undergoing a very significant period of transition to ensure DSA compliance, they would suggest a minimum transition period of 12 months. They 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 (TikTok).

Meta believes 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 they 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 (Meta).