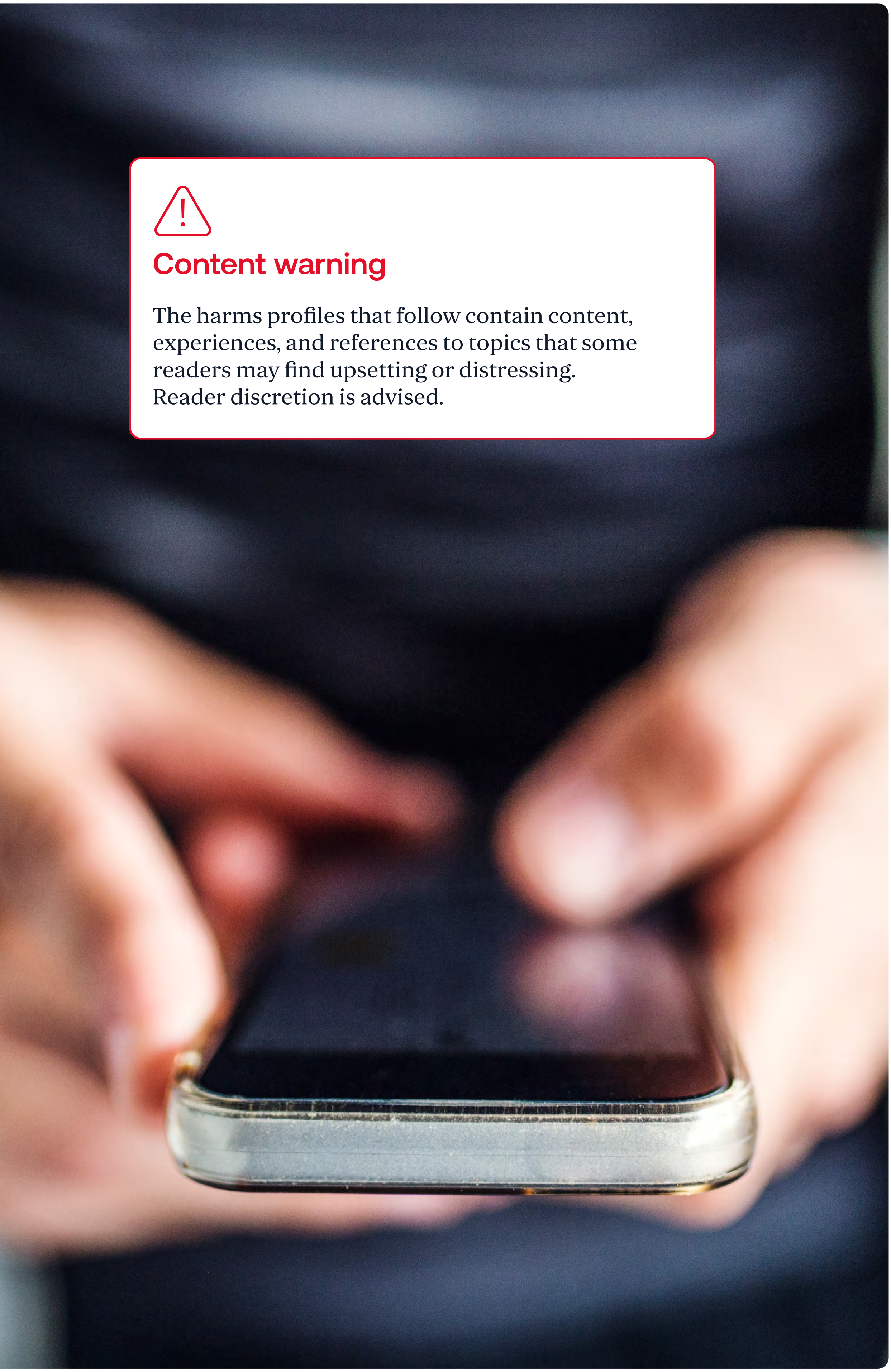
Video-Sharing Platform Services Online Harms Evidence Review

Provided to inform Coimisiún na Meán's approach to VSPS regulation
September 2023

Bringing Ingenuity to Life.
paconsulting.com



Content warning

The harms profiles that follow contain content, experiences, and references to topics that some readers may find upsetting or distressing. Reader discretion is advised.

Contents

1 | Document purpose 6

2 | Scope 8

3 | Methodology 12

4 | Structure 14

5 | Online harm profiles 16

5.1 | Online harm prevalence in Ireland 17

5.2 | Use of video-sharing platforms (VSPs) 19

5.3 | Online content by which a person bullies or humiliates another person 21

5.4 | Online content by which a person promotes or encourages behaviour that characterises a feeding or eating disorder 27

5.5 | Online content by which a person promotes or encourages self-harm or suicide, or makes available knowledge of relevant methods 34

5.6 | Other online content which impairs the physical, mental, or moral development of minors 42

5.7 | Online content by which a person incites hatred or violence 50

5.8 | Offences relating to the online identification of victims, suspects, or vulnerable people 61

5.9 | Online content associated with terrorism 63

5.10 | Online content associated with child sexual abuse 69

5.11 | Online content by which a person's behaviour constitutes harassment or harmful communication 74

5.12 | Online content associated audio-visual commercial communications 82

5.13 | General features of a VSP that can enable harm 88

6 | General response measures 90

7 | Outstanding matters 94

8 | Further areas of potential interest 124

Annex A | Further information on user rights 126

Glossary 130

Endnotes 134

This report is a literature review of available evidence pertaining to online harms on videosharing platforms. It is provided by PA Consulting as independent, expert advice, and one of the information sources intended to inform the development of the Online Safety Code that will apply to video-sharing platforms in the regulatory scope of Coimisiún na Meán.

Coimisiún na Meán was established in March 2023 upon the enactment of the Online Safety and Media Regulation Act 2022. The Online Safety and Media Regulation Act 2022 amended the Broadcasting Act 2009 to establish An Coimisiún and dissolve the Broadcasting Authority of Ireland (BAI).

In addition to undertaking the functions of the BAI as the broadcasting regulator, An Coimisiún will establish a regulatory framework for online safety. This involves issuing a binding Online Safety Code for platforms that allow people to share and view video content on the internet. In Ireland, watching video clips is the most popular activity for children. YouTube, Snapchat, Instagram, TikTok, and Facebook were the services most commonly used by Irish children in 2021.[1]

In developing the Online Safety Code for video-sharing platforms, An Coimisiún must have regard to available evidence pertaining to online harms, and specifically the matters set out in Section 139M of the Online Safety and Media Regulation Act. This report is one of a number of inputs intended to help An Coimisiún to reach an informed consideration of these matters.

The report broadly divides into two sections. The first part comprises a set of ‘harm profiles’, which summarise available evidence about each of the harms in the scope of the Online Safety and Media Regulation Act 2022 (OSMR), and the Audio-Visual Media Services Directive (AVMSD). A preliminary comparative analysis of these legal frameworks generated a list of four common categories of harmful content, which were further sub-divided into 10 research topics (see Figure 1). The second part of this report summarises evidence relating to outstanding Section 139M matters (see Table 22).

Finally, the report outlines further areas of potential interest where research is particularly nascent. An Coimisiún may want to periodically review emerging evidence in these areas, and/or commission new information gathering exercises, to inform future iterations of the Online Safety Code or increase the general awareness of issues in relation to harmful online content.

The report primarily draws on evidence pertaining to Irish VSP usage and experiences, but where data is limited, relevant information from the UK, other EU countries, and the wider world has also been incorporated. Sources published within the past five years have been preferred over earlier publications, although where recent insight is limited, scope has been broadened to include older sources. Minimal information is included on the impact of emerging technologies such as generative artificial intelligence (AI) and the metaverse, because at the time of writing, research into these areas is newly underway.

Report disclaimer

This report has been prepared by PA Consulting Group on the basis of information supplied by the client, third parties (if appropriate), and that which is available in the public domain. No representation or warranty is given as to the achievability or reasonableness of future projections, the assumptions underlying them, or the targets, valuations, opinions, prospects, and returns (if any), which have not been independently verified. Except where otherwise stated, the report speaks as at the date indicated within the report.

1 Document purpose

This report is a review of available evidence pertaining to online harms and online safety, and is provided as independent, expert advice to inform Coimisiún na Meán's approach to developing its binding Online Safety Code.

Specifically, this report contains information to assist An Coimisiún in considering the matters set out in Section 139M of the Online Safety and Media Regulation Act, which must inform the approach to developing the Code and, secondarily, applying it to designated online service providers.

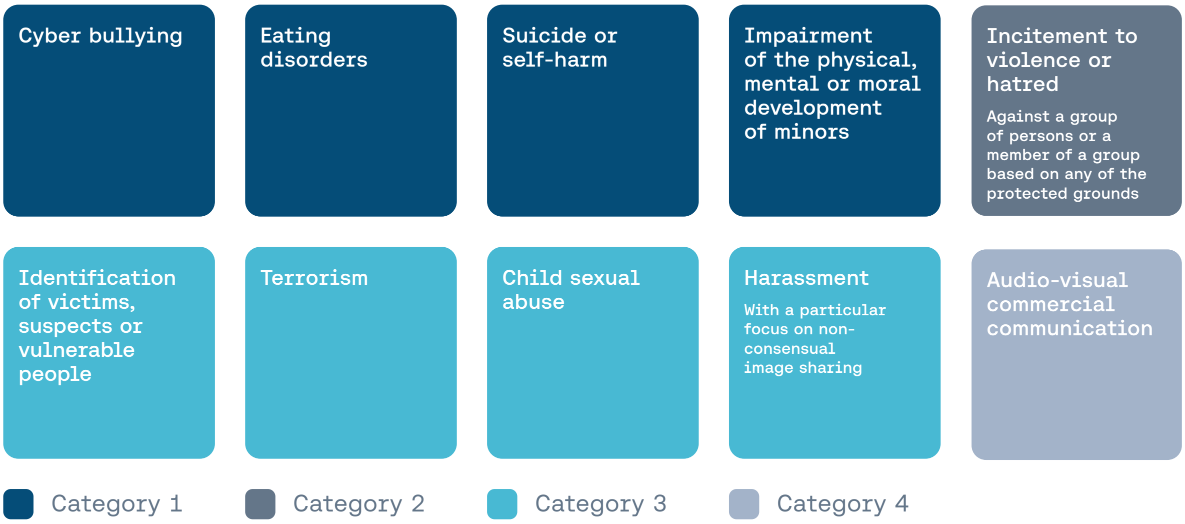
2 Scope

The scope of the research report as follows:

- **A focus on online harms in the scope of the Online Safety and Media Regulation Act 2022 (OSMR) and the Audio-Visual Media Services Directive (AVMSD):** A preliminary comparative analysis of these legal frameworks resulted in offences and specific harms outlined in the OSMR and AVMSD being grouped based on their similarities. This analysis generated a list of four common categories of harmful content, which have been further sub-divided into 10 research topics. Content categories are described in Table 1 (note these are reflected in An Coimisiún's call for inputs). Figure 1 overleaf provides an overview of research topics, with the content categories overlaid.
- **Focus on harms manifesting on Video-Sharing Platforms (VSPs):** This is primarily video content, but also considering other types of content that may be present on VSPs and cause or contribute to harm (for example, image or text posts under videos).
- **Evidence that pertains to Irish users where possible:** If available data in Ireland is limited, research has been broadened out to include evidence from the UK, other EU countries, and the wider world.
- **Sources published within the past five years have been preferred over earlier publications:** Where recent insight is limited, research scope has been broadened to include older sources as well.
- **Emerging technologies (such as generative AI or the metaverse) will have an impact on online harms, but current research on this is limited:** Minimal references are made towards the impact of emerging technologies and the potential harms associated with them.

Content category	Description
1	Content that might impair the physical, mental, or moral development of minors, including age-inappropriate content. This includes the power to protect minors from harmful videos by which: • a person bullies or humiliates another person; • a person promotes or encourages behaviour that characterises a feeding or eating disorder; • a person promotes or encourages self-harm or suicide; • a person makes available knowledge of methods of self-harm or suicide.
2	Content that incites violence or hatred directed against a group of persons or a member of a group based on any of the grounds referred to in Article 21 of the European Charter of Fundamental Rights. The grounds referred to in Article 21 of the Charter include sex; race; colour; ethnic or social origin; genetic features; language; religion or belief; political or any other opinion; membership of a national minority; property; birth; disability; age; and sexual orientation.
3	Content dissemination that constitutes a criminal offence. This includes content that is a public provocation to commit a terrorist offence; offences concerning child sexual exploitation and abuse; offences concerning racism and xenophobia; and online content contrary to Irish law in these and other areas. Schedule 3 of the OSMR provides a full list.
4	Harmful video advertisements (‘commercial communications’). This includes requirements that commercial communications are transparent to users; do not include content that will cause harm to the physical, mental, or moral development of minors; and meet standards in terms of human dignity and non-discrimination. Restrictions on certain products and services are also included, such as alcohol, cigarettes, e-cigarettes, and medicines.

Table 1: Content category descriptions



Cyber bullying

Eating disorders

Suicide or self-harm

Impairment of the physical, mental or moral development of minors

Incitement to violence or hatred
Against a group of persons or a member of a group based on any of the protected grounds

Identification of victims, suspects or vulnerable people

Terrorism

Child sexual abuse

Harassment
With a particular focus on non-consensual image sharing

Audio-visual commercial communication

■ Category 1 ■ Category 2 ■ Category 3 ■ Category 4

Figure 1: The 10 topics resulting from an analysis of the OSMR and AVMSD, mapped to the content categories described in Table 1

> The principal purpose of the [video-sharing platform] service or a dissociable section thereof, or an essential functionality of the service, is devoted to providing programmes, or user-generated videos to the general public [...] to inform, entertain or educate, by means of electronic communication networks and the organisation of which is determined by the video-sharing platform provider, including by automatic means or algorithms in particular by displaying, tagging, and sequencing.

AVMS Directive EU 2018/1808[2]

3 Methodology

This report is a meta study that distils and analyses findings from multiple Irish, European, and international studies and publications. It aims to offer a balanced assessment while recognising the limitations of available data on some aspects of online harms, and signposting divergent views where they exist. Primary data is not included within this report.

This report includes references throughout to examples of harms manifesting on specific platforms. These are included for illustrative purposes only, and are not intended to be exhaustive.

This report has been compiled through an extensive literature review of publicly available information and research, supplemented by a review of information provided by An Coimisiún for the express purpose of informing this report. The contents of the report also draw on PA's expertise in Online Safety.

PA's Online Safety Expertise

PA works with UK and international governments, law enforcement agencies, regulators, and third sector partners (for example, the Internet Watch Foundation and the UK National Society for the Prevention of Cruelty to Children – NSPCC) to understand how online harms are evolving globally (considering offender and victim behaviours, technology developments, and socio-economic factors).

PA has been involved with online harms regulation since 2018 and was the only consultancy to respond to the UK Government's first consultation on the topic. Since then, PA has continued to work with a range of UK public sector entities to prepare for online safety regulation, including with Ofcom on its 'A-SPARC' model to inform the regulation of VSPs, and the National Crime Agency.

PA has extensive expertise on the topic of children's online safety, which has developed and grown since facilitating the first series of multi-sector workshops convened by the UK government in 2014 to come up with potential solutions to improve children's online safety. PA is also a founding industry partner of the **WeProtect Global Alliance** and has developed all three editions of the Global Threat Assessment of child sexual exploitation and abuse online (in 2018, 2019, and 2021). At the time of writing, PA is leading the development of the 2023 edition of the report.

Through work across the defence and security sectors, PA has a complementary, in-depth understanding of other high threat harms such as terrorism, violent extremism and hate speech, as well as of the methods, technologies, and tools used by the platforms who host user-generated content to detect, assess, block, and remove harmful material.

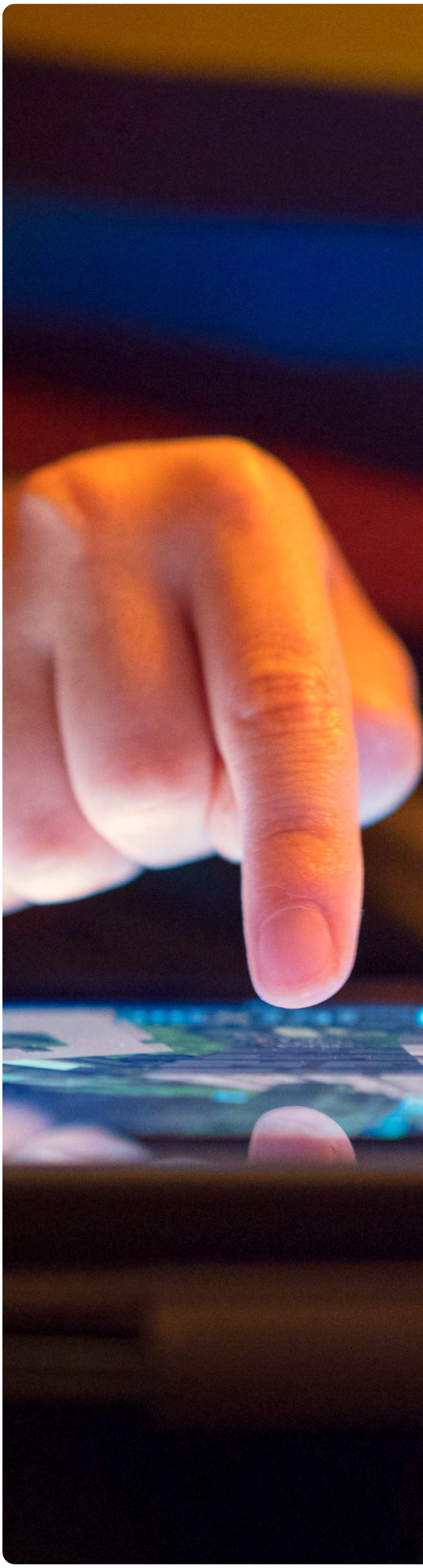
The methodology used for desk-based research was as follows:

1. Define scope
   The scope of the research exercise was agreed with An Coimisiún, informed by a review of the matters set out in Section 139M of the Online Safety and Media Regulation Act.
2. Define research topics and timelines
   The scope of the research exercise was broken down into smaller research topics (see Figure 1), and the outstanding Section 139M matters. A timeline was defined, stipulating iterative research and writing periods.
3. Agree sources
   A desk-based literature review of publicly available information was conducted, complemented by analysis of information provided by An Coimisiún and PA's own online safety expertise.
4. Iteratively conduct research, summarise findings, and compile references
5. Review and refinement.

The report was taken through PA's internal quality assurance process, involving review by expert SMEs, and feedback incorporated prior to finalisation.

4

Structure



The report broadly divides into two sections:

1. Part 1 (Section 5) comprises 11 harms profiles. These are individual summaries of current literature specific to the online harms potentially in regulatory scope (see Figure 1). Section 5 contains a more detailed description of how the harms profiles are structured.
2. Part 2 (Section 7 onwards) includes a summary of research specific to outstanding matters set out in Section 139M, which are not covered within the harms profiles.

5 Online harms profiles

5.1 Online harm prevalence in Ireland

This report has, where possible, considered literature that draws on Irish statistics, research, or case studies. The purpose of this section is to provide a high-level summary of the prevalence of online harm in Ireland, based on the findings of the research.

In Ireland, similar to the rest of the developed world, watching video clips is the most popular activity for children. YouTube, Snapchat, Instagram, TikTok, and Facebook were the online services most commonly used by Irish children in 2021.[3]

Within this context of prolific VSP use, particularly by children, this review examined online harms potentially in regulatory scope. Included below is prevalence data specific to the online harms that were in the research scope:

Bullying: Bullying is affecting Irish children's ability to learn and feel safe at school; 53 percent of 8–12 year-olds (in June 2022[4]) and 11 percent of 9–17 year-olds (between December 2019 and October 2020[5]) reported experiencing bullying online. This is supported by the findings of the Cybersafe Kids survey (2021–2022), in which 28 percent of Irish children reported having experienced a form of online bullying, with exclusion from a chat or messaging group being the most prominent.[6] Harm statistics are affected by underreporting, with some studies citing that the majority of children would never tell their parents if they were cyberbullied or wouldn't know how to.[7]

Eating disorders: Research conducted on 9-17 year olds by Ireland's National Advisory Council for Online Safety (NACOS) between December 2019 and October 2020, revealed that 26 percent of children surveyed had seen harmful online content in the previous 12 months. This survey found that 11 percent of 9–17 year-olds in Ireland have been exposed to content or discussions on 'ways to be thin'.[8]

Self-harm and suicide: The NACOS survey also concluded that 13 percent of 9-17 year olds surveyed had been exposed to self-harm content and nine percent had been exposed to content which detailed suicide methods.[9]

Gratuitous violence: Unnecessarily violent or gory content is the second most prevalent form of harmful content Irish children are exposed to online, with 18 percent of children who have experienced harmful content matching it to this category. However, older children (13–17 years old) are at greater risk of exposure (>25 percent).[10]

Sexually explicit: The Irish Society for the Prevention of Cruelty to Children (ISPCC) and Children at Risk in Ireland (CARI) have endorsed the findings of a study which found that sexual violence commonly seen in pornography was also found in half of police interview transcripts for sex abuse cases committed by a child against another child (child-on-child).[11] Almost a fifth of all children aged 9–17 had seen sexually explicit content in the past year.[12]

Harassment and abuse: One in five young women aged 19–25 in Ireland has suffered intimate relationship abuse. Of these, 49 percent experienced online abuse facilitated by digital technology.[13] Nearly a third (30 percent) of young women aged 16–29 said they had experienced cyber harassment in the past five years.[14]

Child sexual abuse (CSA): As of December 2022, Ireland's national centre for combatting illegal content online (Hotline.ie) had removed 14,772 pieces of child sexual abuse material (CSAM), which is 25 percent more than in the previous 21 years combined.[15]

Advertisements: The Advertising Standards Authority for Ireland has found that almost half of influencer adverts are not correctly tagged as advertising, and a quarter of Irish consumers who have purchased a product as a result of an influencer promoting it subsequently felt misled.[16]

30%

of young women aged 16–29 said they had experienced cyber harassment in the past five years.

5.2 Use of video-sharing platforms (VSPs)

VSPs are a type of online service primarily used to upload and share video content.

They are widely used by a broad range of internet users, particularly young people.

- In the academic year 2021–2022, 95 percent of 4,408 Irish children aged 8–12 stated that they owned a smart device. Smart device ownership also increases by age, from 89 percent of eight year-olds to 99 percent of 12 year-olds.[17]
- Watching video clips is the most popular online activity for Irish children, determined by a study of 765 children aged 9–17 in 2021[18] and reiterated in 2020/21 study of 4,408 Irish 8–12 year-olds.[19]
- Use of VSPs was the most cited activity among all children aged 3–17 (95 percent) in a UK study of 6,600 children in 2021, with 31 percent posting content they had made themselves.[20] This is supported by the findings of the Cybersafe Kids survey, which found that of the Irish children aged 8–12 years who participated, 27 percent posted videos of themselves online, with the most common destination of those videos being TikTok (74 percent).[21]
- Watching online videos is also the favourite media activity among American 8–18 year-olds (62 percent of 1,306 US children listed it as their favourite).[22]
- YouTube (51 percent), Snapchat (43 percent), Instagram (26 percent), TikTok (22 percent), and Facebook (11 percent) were the top five online services (including non-VSPs) used by Irish children aged 9–17 in 2021.[23]

There is minimal additional data regarding VSP usage by Irish audiences, but a recent study undertaken by the UK by the Office of Communications (Ofcom) provides potentially relevant insight. According to this research, although most users engage passively with VSP content (for example, 74 percent browsing or scrolling content, and 53 percent searching for specific videos/content), a significant percentage (45 percent) actively share content with friends, family, or other platforms, and 32 percent post or upload content.[24] The 2023 U.S. Surgeon's General's Advisory report found that children aged 12–15 who spend more than three hours per day on social media are twice as likely to experience mental health issues such as depression and anxiety.[25]

According to Ofcom's research, the majority (70 percent) of those who use VSPs had seen or experienced something potentially harmful in the past three months.[26] This is somewhat representative of the experiences of Irish children, as according to the Cybersafe Kids 2021–2022 survey, 26 percent of 4,408 aged 8–12 had seen or experienced something online that bothered them, with this proportion being significantly higher for eight year-olds (35 percent) than for the other age groups (24–28 percent).[27]

Ofcom's research also revealed that young people are far more likely to actively engage with VSPs.[28] This is significant as active engagement exposes them to different types and severity of harm through contact with other (potentially adult or malicious) users.[29]

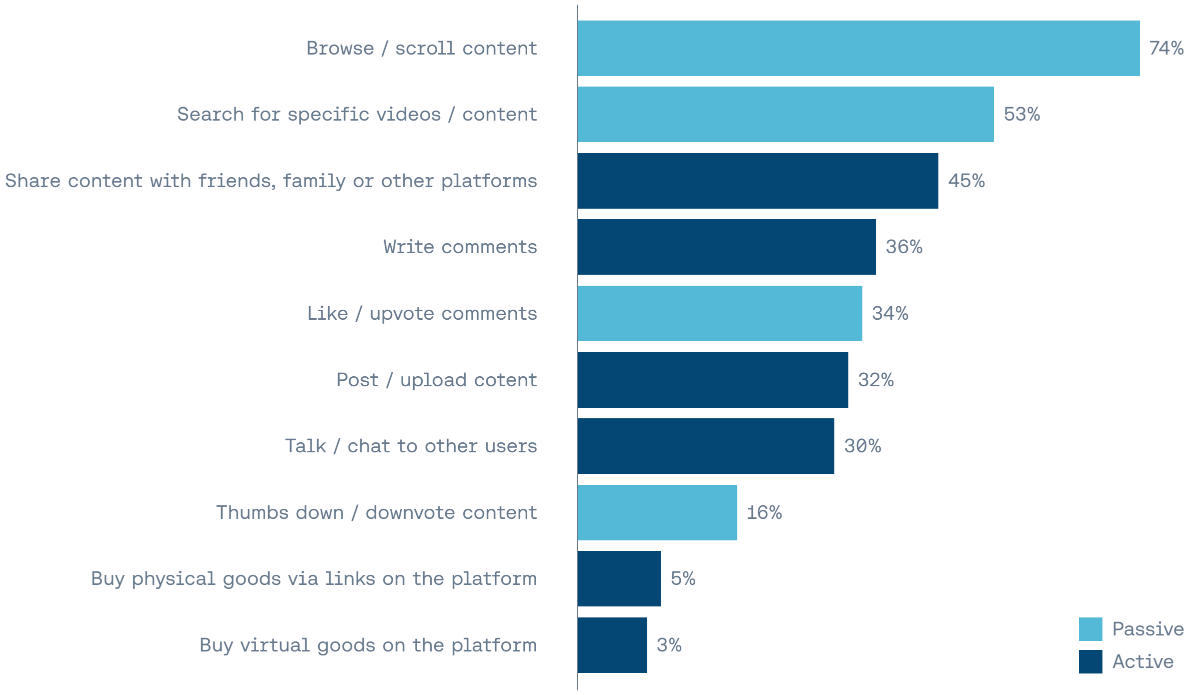


Figure 2: Q: In general, what do you tend to use VSP services for? N=1,980[30]

This section of the report provides detailed information pertaining to each of the harms in the scope of the Online Safety and Media Regulation Act 2022 and the Audiovisual Media Services Directive 2018, and specifically how they manifest on VSPs. Each harm profile is structured as follows:

1. **Description**
   Definition of the harm, contextual information, user perspectives, and examples of how this harm manifests on VSPs.
2. **Prevalence and risk**
   Key statistics and evidence regarding the availability of the harmful content online, risks of exposure, and additional harm specific information.
3. **Impact**
   Summary of available evidence about how this harm can directly impact users or lead to further indirect impact. Where possible, additional data on impact rates is included, alongside case studies providing qualitative evidence of that impact.
4. **Enabling VSP features**
   Summary of 'risk vectors': VSP features that can cause or contribute to propagate the harm in question.
5. **Specific VSP response measures**
   Summary of response measures that can be implemented to address the harm.

5.3 Online content by which a person bullies or humiliates another person

This section will focus on the prevalence and impact of content related to cyber bullying on VSPs. Additionally, it will cover the VSP features that can enable this harm, as well as any specific response measures that can mitigate it.

Cyber bullying is bullying with the use of digital technologies. It can take place on social media, messaging platforms, gaming platforms, and mobile phones.

It is repeated behaviour aimed at **scaring**, **angering**, or **shaming** those who are targeted.[31] Unlike bullying offline, online bullying can follow users wherever they go via their devices.

> I am being bullied by a girl at school. She has taken photos of me and posted them on Snapchat calling me fat and ugly and how I will never have a boyfriend. I have been having suicidal thoughts as this girl is really popular and she has turned my whole year against me.
>
> Girl, aged 14[32]

Increased use of social media and digital platforms among children is leading to new social interactions and changing the nature and reach of bullying activities impacting children. The mobility and pervasiveness of the tools young people use to conduct their social life is greater than ever. For young people, offline and online are indistinct.

Bullying, hate speech, and hurtful messages can now follow a child into the most private corners of their life. Digital interactions can reach far more bystanders – transforming private hurt into public humiliation. Editing and filtering technologies means the range of hurtful communications are bound only by imagination.

Richardson, Milovidov and Blamire, Bullying: Perspectives, practice and insights. Council of Europe. (2017); paraphrased.

Receiving abusive comments about a person's appearance	Sexual bullying	Telling another to kill themselves
Pressuring another into sharing sexual images	Threatening another with images being posted online	Cyber stalking
Trolling*	Exclusion	Shaming
Reputational attacks	Extortion/sextortion	Enticing or goading persons online to self-harm, commit a crime, or dangerous act
Doxing*	Flaming*	Harassment

Table 2: Examples of how cyber bullying manifests on VSPs[33, 34]

*See glossary[35,36,37]

Consider a hypothetical yet common example of 'Steven', aged 14, who posts a video of himself dancing on a social media platform. He loves to dance, even though he doesn't consider himself to be good at it, but he's trying to improve. He is not very popular; he is very insecure and does not have a strong support group among his peers. Yet, to his credit, Steven tries to get out of his comfort zone and explore who he could become by posting the video.

A lack of a sense of dignity and self-worth might be motivating his actions too. Subsequently, he gets laughed at and humiliated in comments by his peers; someone remixes his video into a derisive meme which now seems to go viral; response videos mocking him are created by other peers.

Milosevic, Changing the Paradigm for Cyberbullying Intervention and Prevention: Considering Dignity, Values, and Children's Rights (2021) Available at: https://www.ispcc.ie/guest-post-changing-the-paradigm-for-cyberbullying-intervention-and-prevention-consideringdignity-values-and-childrens-rights/

23%

of children experienced bullying online between 2017 and 2019 (on average across 19 European countries)

5.3.1 Prevalence and risk of harm

- **Public perception:** Bullying is the greatest concern among a representative sample of 1,000 parents and 2,000 children in the UK when considering the harms children face online.[38]
- **Prevalence in IE & EU:** In Ireland, based on a 2021 survey, 11 percent of 9–17 year-old internet-using children experienced bullying online between December 2018 and October 2019,[39] but this figure rises to 23 percent of 21,964 children when considering an average across 19 European countries surveyed between 2017 and 2019.[40] Another study conducted in Ireland of 340 children aged 8–12 found that 53 percent of respondents had been cyber bullied, and 18 percent had been cyberbullied in a way that really affected their ability to learn and feel safe at school.[41]
- **Prevalence on VSPs:** Based on a 2021 survey of 1,958 people in the UK, 26 percent of VSP users declared that they had been exposed to bullying, abusive behaviour, or threats while using VSPs in the past three months. In the same survey, 36 percent had experienced trolling in the past three months.[42] Furthermore, 22 percent of 1,500 children in the UK said someone has posted an image or video to bully them.[43]
- **Risk of harm:** Of all the harms surveyed across 2,000 UK children, experiencing bullying from a person they know was found to have the greatest prevalence of high affect impact on the children (of the 15 percent of children who experienced it, 64 percent reported a high affect impact).[44]
- **Under reporting:** Based on a 2020 survey of children aged between 10 and 15 in the UK, more than half (52 percent) of children who experienced online bullying behaviours said they would not describe these behaviours as bullying, and one in four (26 percent) did not report their experiences to anyone.[45] Furthermore, in Ireland, a survey conducted in 2022 of 340 children aged 8–12 found that 60 percent of respondents would never tell their parents if they were cyberbullied, or wouldn't know how to have this conversation.[46] Despite this, in 2022 and 2021, cyber bullying was the most frequently reported problem to European helplines (14 percent of 56,891 reports).[47]

5.3.2 Impact of this harm

The impact of this harm resembles the impact from offline bullying. Both online and offline bullying can result in significant harm to the physical, mental, and moral development of children. The table below provides some examples of these impacts.

	Physical	Mental	Moral
Direct impacts	• Weight loss • Self-harm • Stomach pains • Sleep problems • Headaches • Tension • Bedwetting • Fatigue • Poor appetite	• Anxiety • A feeling of inadequacy • Low self-esteem • Isolation • Personality change	• Increased device privacy • Reduction in academic achievement
Indirect impacts	• Substance abuse • Suicide	• Higher risk of psychosomatic problems (illnesses caused by anxiety or worry) • Lonely children are twice as likely to be groomed online	• School truancy

Table 3: Impacts of cyber bullying on the physical, mental, and moral development of children[48, 49, 50]

Impacts sourced from (unless noted otherwise): Kowalski et al, Psychological, physical, and academic correlates of cyberbullying and traditional bullying, Journal of Adolescent Health. (2013) Available at: https://www.sciencedirect.com/science/article/pii/S1054139X12004132

After an argument with Connie (26–30), Connie's friend created a false social media profile under the name 'Joe King'. The account had no profile picture or other personal information attached. The account was used by the friend to spread malicious and false rumours about Connie in the local area amongst friends and her places of work. Connie believes that the harassment would not have been carried out if her friend had not been able to make a fake account so quickly and easily, with no fear of consequences.

Revealing Reality, How people are harmed online: Testing a model from a user perspective (2022) Available at: https://www.ofcom.org.uk/__data/assets/pdf_file/0023/244238/How-people-are-harmed-online-testing-a-model-from-auser-perspective.pdf

5.3.3 VSP features which can enable harm

Design features, typical use cases, and community types can further enable the risk of harm occurring on VSPs. Below are examples of this occurring:

Anonymity: Anonymity is the absence of personally identifying information. It may embolden users to abuse others and encourages deception, such as hiding a child's activity from their parents. Within VSPs there are several features relevant to the anonymity of users, listed below.[51]

1. Pseudonymised usernames which do not reflect a user's legal name. This can have a disinhibiting effect on users that can lead to harmful behaviours such as bullying or trolling.
2. Identification verification during account registration. Requested details are often not verified, making it difficult to trace the real person behind the account.
3. Alt accounts, or multiple accounts on the same service, which limit parental oversight and bypasses parental safeguards placed onto child accounts. It can also enable instances of digital self-harm or self-bullying.[52]
4. Social bots which emulate human communication and can be co-opted for co-ordinated harmful contact or to spread harmful content at scale.
5. Embedded anonymity features that can be added to mainstream platforms. For example, a child can ask friends to ask or answer questions anonymously using add-on features from a third-party app.
6. Randomised meets or services that offer video chats between randomised and anonymised individuals for real-time interaction. These situations can increase the risk of contact with malign actors.
7. Defined networks which used by some services to allow users within a defined network (such as those with a school email or geographical tag) to post anonymous messages. The localised nature of these networks can create a breeding ground for bullying.

Disinformation: This intentionally false or misleading information can take many forms, from memes to low-quality clickbait.[53] VSP features affect the distribution of disinformation and the impact this has on bullying.

1. Transient/disappearing content that expires after a certain amount of time and encourages users to share in the moment. These posts often disappear before they can be moderated, and can be difficult to report.
2. Easily accessible contact lists or groups of contacts make the wide sharing of content seamless. Private messaging channels can facilitate the rapid spreading of content containing disinformation.

Recommendation systems: These systems are used for item/content/network filtering based on user preferences and/or past behaviour.[54] Within VSPs, features associated with bullying and recommendation systems include:

1. Connection recommendations – adults or malign actors can adopt similar interests to young people so that they are introduced to young or vulnerable users.
2. Number of friends – visual popularity metrics such as friends, followers, or subscribers encourage young people to add strangers to improve their social status.
3. Tagging – tagging others' usernames or creating hashtags can be done with the intention of scaring, angering or shaming other users.[55] Tagging can also be used to aggregate content of a similar nature (or allow for ease of searching), potentially increasing the reach of content used for cyber bullying.

An anti-bullying expert has urged adults to examine how they use social media, as the video of an unprovoked assault on a teenager in Ireland received over five million views.

The majority of views are due to the video being shared by adults on social media, according to Unesco chair on bullying and cyber-bullying, and director of Dublin City University's anti-bullying centre Professor James O'Higgins Norman.

Fiona Jennings of the ISPCC added that people often people come across something online which shocks them and in order perhaps to share their empathy or to empathise with those particular things they share them.

"But in fact, that's actually feeding the algorithm that promotes and amplifies this content, which in turn actually brings it to a wider audience," she told RTÉ's Morning Ireland.

https://www.irishexaminer.com/news/arid-41142553.html

Live-streaming: This technology lets users watch, create, and share videos in real-time. All users need to be able to live-stream is an internet enabled device, such as a smartphone or tablet, and a platform (such as a website or app) to live-stream from.[56] VSPs features associated with bullying and livestreaming include:

1. Public profiles – VSPs often set live-streams to public by default, making them visible to millions of users.
2. Direct messaging – enabling viewers to move from public interaction to private messaging, providing a private place for harm to occur.
3. Likes – with visualisation design choices such as hearts or other emojis to enable others to exploit the desire for social affirmation, which is particularly strong in young people.
4. Live chat – enabling viewers to interact with the live-streaming of young people in real-time, with six percent of children who have live-streamed being asked to change or remove their clothes on camera.[57]
5. Moderation – in recent interviews by Ofcom, platforms say that live-streaming and other ephemeral content presents moderation challenges that are distinct from other types of content.[58]

5.3.4 Specific response measures

Quantitative research on social solutions (solutions that are not reliant on technological interventions) is sparse. However, a 2016 EU study on media literacy identified 547 media literacy projects (including subjects on online behaviour), with themes including frontline support by highlighting resources (32 percent of projects); end user engagement (20 percent); improving skills in critical thinking (74 percent); and media use (70 percent).[59]

Toolkits for children, teachers, educators, or parents and caregivers: Due to the similarities between online and offline bullying, research suggests that cyber bullying training should be administered within already established risk-prevention programmes for bullying.[60] This is further confirmation of the indistinct nature between online and offline environments.

A report on cyber bullying prevention and intervention states: "The impacts from cyber bullying arise from how young people view themselves in relation with others, as well as how they see and treat each other, based on what they believe about how to attain worth and success. Viewing cyber bullying merely as a matter of tech features, online behaviour or an 'online safety' issue severely limits our thinking in terms of finding solutions to this problem."[61] It is therefore important to consider social solutions as significant in combatting cyber bullying.

Much of the existing response measures for cyber bullying are classified as social solutions; with various toolkits in existence for children, teachers or educators, and parents or caregivers.[62] These toolkits cover a range of topics, but several relate specifically to:

- Cyber bullying
- Sexting
- Peer pressure
- Self-esteem
- Live-streaming
- Permission and consent to sharing media content
- Identifying negative behaviours.

Within Ireland, there are several NGOs that provide preventative and reactive advice as well as support for cyber bullying:

- www.ispcc.ie
- www.npc.ie
- www.spunout.ie
- www.webwise.ie
- www.barnardos.ie
- www.watchyourspace.ie
- www.familyfriendlyhq.ie
- www.antibullyingcentre.ie
- www.cybersafekids.ie

The Department of Education has developed a new action plan (Cinealtas: Action Plan on Bullying) to prevent and tackle bullying in schools.[63] This is based on four key principles:

- Prevention: Empathy and education that provide the foundation for inclusion and respect, among other things.
- Support: Tangible and targeted support that provide a framework for school communities to collaborate.
- Oversight: Visible leadership that creates positive environments for young people.
- Community: Building inclusive school communities which support positive relationships.

Further, the Department of Education also co-funded Webwise.ie, a source of information, advice, and education for young people, teachers, and parents covering a range of internet safety topics.

FUSE, a programme run by the Anti-Bullying Centre designed to comply to UNESCO's Whole Education approach to bullying, has had significant impact on cyber bullying measures in Ireland since 2019.

Of the 8,842 primary school students registered:

- 93 percent are more confident in knowing who to tell if something bothers them online
- 93 percent are more confident in knowing how to use social media safely
- 98 percent are more confident in knowing not to share personal information online
- 99 percent are more confident in knowing how to respect others online.[64]

Technology solutions: Notwithstanding the high reliance on social solutions, there are also 'safety technology' interventions that exists in response to cyber bullying. For example, an ongoing Irish project (Open Standards for AI-based Bullying Interventions[65]) aims to create standards for proactive antibullying interventions on social media platforms by soliciting children's feedback about the effectiveness of these interventions and their impact on children's rights to safety, privacy, and freedom of expression.

There are further general responses defined in Table 21.

5.4 Online content by which a person promotes or encourages behaviour that characterises a feeding or eating disorder

This section will focus on the prevalence and impact of content related to eating disorders on VSPs. Additionally, it will cover the VSP features that can enable this harm, as well as any specific response measures that can mitigate it.

Online content often referred to as 'pro-anorexia', 'pro-ana', and 'pro-bulimia', or 'pro-mia' promotes the harmful behaviour and mindset that forms part of eating disorders.[66]

The influence of social media can contribute to the development of disordered eating. Pro-eating disorder websites and other online communities portray eating disorders as positive and promote harmful weight control practices.[67]

“

A 14 year-old girl got in touch with the helpline to discuss her engagement with pro-anorexia and self-harm material. She had uploaded a full-body photo of herself (with clothes on), which was then shared in a pro-anorexia group. She shared that she felt good receiving likes and praise, and that she has never felt happier about her body. She wondered why aspiring to be very thin is seen as a bad thing. She also discussed self-harming, being proud of her scars, and watching suicidal ideation videos online.

Girl, aged 14

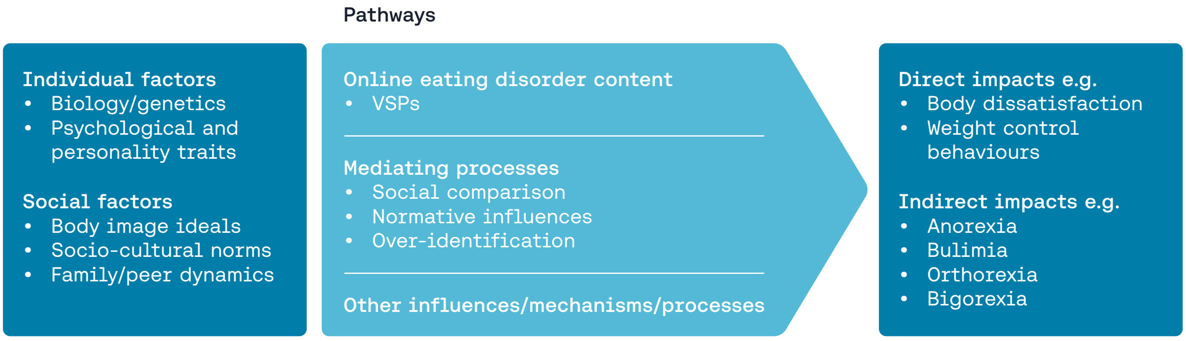
Better Internet for Kids, Classifying and responding to online risk to children (2023)

For some minors, the depreciation of existing psychological difficulties or the normalisation of pathological behaviour is enabled by their digital experiences. This effect is compounded by some minors finding it difficult to self-regulate their digital engagement.[68]

This is particularly apparent when considering eating disorders, with digital engagement typically taking the form of three activities: comparing; curating; and community. This digital engagement on online platforms can provide refuge for those with a lived experience of eating disorders, but equally, online content could also trigger and prolong harmful behaviour.[69]

Disordered eating attitudes are linked to self-esteem, body image, body dissatisfaction, and the use of social media: There is a well-established link between the development of eating disorders and body dissatisfaction.[70, 71] A significant factor in body dissatisfaction is social media, where unrealistic beauty ideals are popularised.[72] This characterises the risk of cumulative exposure to content portraying 'ideal' or unrealistic body images, as opposed to pro-eating disorder content.

The latest research, presented by Khatwa et al., suggests there is a complex causal pathway demonstrating how the interaction of multiple causal factors can lead to the development of eating disorders; and how online eating disorder content can influence and interact with pre-existing individual and social factors to impact young people's body image and disordered eating, as shown below.[73]



Pathways

Individual factors
- Biology/genetics
- Psychological and personality traits

Social factors
- Body image ideals
- Socio-cultural norms
- Family/peer dynamics

Online eating disorder content
- VSPs

Mediating processes
- Social comparison
- Normative influences
- Over-identification

Other influences/mechanisms/processes

Direct impacts e.g.
- Body dissatisfaction
- Weight control behaviours

Indirect impacts e.g.
- Anorexia
- Bulimia
- Orthorexia
- Bigorexia

'Thinspiration' material (for example, quotes)	Eating disorders as a lifestyle choice	'How-to' guidance
Comparisons (for example, portion size, body image, stories of achievement, weight loss accounts)	Peer pressure (for example, competitive behaviour, extreme target setting, insults, or criticisms)	Positive endorsements on content depicting persons in a disordered state
Peer acceptance (for example, a sense of belonging)	Support strategies (for example, how to hide eating disorder symptoms, suppress appetites, or otherwise bring about weight loss)	Counter-culture narratives

Table 4: Examples of how the promotion or encouragement of eating disorders manifests on VSPs[74, 75]

5.4.1 Prevalence and risk of harm

- Prevalence: The number of children aged 12 to 16 who see ways to be very thin on the internet at least every month or more often varies across countries, ranging from three percent (Germany) to 32 percent (Poland). The average was 12 percent across the European countries surveyed.[76] In the same survey, 21 percent of children reported seeing this content a few times over the past year.[77] In Ireland, a survey of 9–17 year-olds found that 11 percent of respondents had been exposed to content or discussions on 'ways to be very thin'.[78]
- Prevalence on VSPs: Based on a 2021 survey of 1,958 people in the UK, 21 percent of VSP users declared that they had been exposed to negative body image, excessive dieting, or harmful eating disorder content while using VSPs in the past three months. However, this rose to 34 percent when considering only the 215 13–17 year-olds in the same survey.[79] In a US study led by the NGO Centre for Countering Digital Hate, researchers found 56 TikTok hashtags hosting eating disorder videos, which collectively had over 13.2 billion views.[80] Australian researchers have also shown that Instagram is host to a network of eating disorder accounts reaching 20 million unique followers on the platform.[81]
- Girls are more likely to be exposed to this content than boys: In the UK, a survey of 13–17 year-olds found that girls are almost twice as likely (23 percent versus 12 percent) to be exposed to content associated with 'ways to be very thin'.[82] A survey of children aged 12–16 in European countries found that 15 percent of girls are exposed to content that promotes ways to be very thin, compared to nine percent of boys.[83] This is further reinforced by data obtained from a survey of 3,656 15–30 year-olds in the US, Finland, Germany, and the UK, which demonstrated that 22.9 percent of females but only 18.7 percent of males were exposed to harm-advocating eating disorder content.[84] However, when exposed, males are equally vulnerable to the potential harms.[85]
- Children are particularly at risk of exposure to this content: In a survey of 214 UK children, 4 percent of 13–17 year-olds stated they had been exposed to negative body image, excessive dieting, or eating disorder promoting content in the past three months. This is compared to only 20 percent of 18–84 year-olds stating the same exposure.[86]

5.4.2 Impact of this harm

Research evidence concludes that social media usage is associated with increased body image concerns and engagement in disordered eating behaviours,[87] and that these relationships strengthen over time with continued use.[88, 89]

In a study of 565 US residents who were 15 years or older and endorsed posting thin-deal or body image content on social media, 61 percent stated that this content elicits negative or bad feelings, and/or lowers self-esteem, with seven percent of responses specifically mentioning anxiety and/or depression. Furthermore, in the same study, 26 percent of respondents felt that this content triggers a desire to engage in eating disorder behaviours and 23 percent felt that it was promoting thinness as attractive or increasing the pressure to be thin.[90]

The impact of social media on eating disorder prevalence may already be occurring in Ireland, as the Health Service Executive's National Clinical Programme for Eating Disorders projects that an estimated 188,895 people in Ireland will have an eating disorder at some point in their life.[91] Furthermore, the Heath Research Board statistics show that eating disorders represented 18 percent of all psychiatric and hospital admissions for under 18s in 2020,[92] demonstrating the already significant impact of eating disorders on young people.

> “
>
> I recently came across this section on Twitter which was all about weight loss and had threads on how to starve yourself. It also had pictures of extreme waists and stuff. This really affected me, to the point that I had to delete the app entirely. Ever since I've been feeling strange about myself and my body.
>
> Girl aged 17, Childline[100]

> Unfortunately, it's very common for my patients to talk about the negative impact of social media on body image and self-esteem. When on these platforms, they are infiltrated with images of their peers who appear to have it ‘all’ – perfect bodies and perfect lives. As most of us know, many images on social media are doctored with filters and editing tools, so the bodies portrayed are unrealistic and unachievable. My patients describe engaging in negative social comparisons with people portrayed on social media, feeling inadequate in comparison to these ‘perfect’ peers, and sometimes end up feeling like a failure.

Dr. Matthews, Clinical Director of the Eating Disorders Program, Cincinnati Children’s Hospital, USA[89]

	Physical	Mental	Moral
Direct impacts	• Weight loss • Purging • Excessive exercise • Binge eating • Fasting	• Body dissatisfaction • Weight control behaviours • Low self-esteem • Cognitive restraint • Worry	• Fear or shame
Indirect impacts	• Other medical complications • Death	• Depression • Anxiety • Anorexia • Bulimia • Orthorexia • Bigorexia • Other eating disorders	• Further impacts associated with ‘doxing’ or the exposure of identity online

Table 5: Impacts of content promoting eating disorders on the physical, mental, and moral development of children[93, 94, 95, 96, 97, 98, 99]

Gabi, 16, reflected that the content she had engaged with when younger, which focused on eating disorders, had felt good at the time but had actually been bad for her in the long-term.

Gabi lives with her parents and younger sibling in Scotland. In her spare time, she likes to see her friends who live all over Scotland, and spends a significant proportion of her time gaming, watching shows, and using social media.

Gabi talked about experiencing several mental health difficulties, including having struggled with an eating disorder. She found the transition from primary to secondary school challenging and found it hard to make friends in her year. She became good friends with a girl several years above her at school who introduced her to 'pro-eating disorder' communities online known as 'pro-ana' groups on a blogging website.

"She [the friend] interacted with a lot of like pro-anorexia content and people who romanticise mental illness. So, it's kind of this thing where like I have depression and you think it makes you cool and mysterious."
Gabi

Having begun to spend time looking through these blogs and engaging with the content and other users, Gabi began to write her own pro-anorexia blog, posting 'low-calorie meal ideas' as well as 'inspirational posts'. She explained that she enjoyed the popularity that her blog received on the site which contributed to her wanting to post more on it.

"I'd go online every night and post these blogs, eating disorder [recipes, products and methods]."
Gabi

Gabi explained that she was posting content from when she started secondary school, aged 11, until she was around 14 years old. Throughout this time, she felt like it was a positive influence on her and enabled her to connect with like-minded people. Now, aged 16, she looks back on this time and realises that these online communities that she was a part of were having a negative impact on her by encouraging and prolonging her eating disorder.

"Anorexia was a huge problem for me for a long time, very much fuelled by these online communities."
Gabi.

Revealing Reality, Research into risk factors that may lead to children to harm online (2021) Available at: https://www.ofcom.org.uk/__data/assets/pdf_file/0021/245163/children-risk-factors-report.pdf

Aria wanted to find connection with people who'd had similar experiences, so she started engaging with content that resulted in her developing extremely negative and self-reinforcing behaviours.

Aria (18–20) is a student. She has a history of mental health issues stemming from her parents' divorce and the three-year development of an undiagnosed chronic disease which led to severe physical and psychological symptoms. From age 13–17, Aria described going through a 'sad' phase where she was an 'emo'. She started an anonymous 'sad' account on a social media platform which she used to post and follow pages related to themes like depression and self-harm to try to find and connect with people who were going through similar things.

Through this page, she was added to an anonymous messaging app group purporting to be 'supporting' people with eating disorders. It was called 'girls ED support group' – and turned out to be an anonymous group of users who were 'coached' in eating disorder tips. The group leaders asked the girls to submit their weight on a weekly basis and, if a girl had gained weight, they would have to send visual evidence of them self-harming as punishment.

Aria was in the group for 1.5 years and became very thin. She left the group when a friend stole her phone and read the messages. Shortly after, she attempted to overdose and was admitted into full-time psychiatric care and remained in care for approximately one year. Aria still has twice weekly therapy sessions and takes medication for her depression.

Revealing Reality, How people are harmed online: Testing a model from a user perspective (2022) Available at: https://www.ofcom.org.uk/__data/assets/pdf_file/0023/244238/How-people-are-harmed-online-testing-a-model-from-auser-perspective.pdf

> Don't be scared to un-follow those who trigger you, because the people you surround yourself with on Instagram have just as much influence as those who you are friends within your day-to-day life.
>
> Sophia, Beat Eating Disorders[112]

5.4.3 VSP features which can enable harm

A study on social media and body image argues that the interactive format and content features of social media, such as the strong peer presence and exchange of a multitude of visual images including video – working via negative social comparisons, transportation, and peer normative processes – can significantly influence body image concerns.[101] VSPs are particularly enabling, as visual platforms have been found to be more dysfunctional for body image than more textual platforms.[102] As a result, online media may have a stronger effect on self-objectification parameters compared to traditional media (such as television).[103]

Visual editing (such as image manipulation): Social media content is frequently filtered and edited by computer software which contribute to unrealistic standards of beauty portrayed on social media. These edits are often undetectable.[104, 105]

Personalised and varied user generated content: Any form of content that has been posted by users on online platforms, tailored or crafted for a user's feed. VSPs offer a rich modality of media, transporting individuals to immersive domains that can encourage suspension of belief and attitude change.[106] The selective curating and sharing of content by users portrays eating disorders as a desirable lifestyle within pro-ana or pro-mia communities.[107] Similarly, digital pruning is the act of selecting who users follow so they can manage what content they see.[108]

Engagement mechanisms: This can include tagging, or using hashtags or other means for seeking out certain content producers or types of content or to provide feedback, critique, support, or share user generated content.[109]

Social media sites cater to communities of like-minded individuals, offering easy and frequent access to similar others. Liking or providing feedback to other users' individual posts or otherwise interacting with existing content, has been shown to perpetuate or normalise harmful behaviour.[110] This can contribute to the competitive nature of eating disorders.[111]

Recommendation systems and algorithms: Algorithms that deliver personalised content to users target content that is often more extreme, less monitored, and designed to maximise user engagement; resulting in the intensification of the association between social media and eating disorders.[113]

Social media influencers: An influencer is a social media user high in social standing who has the power to affect their followers' beliefs and purchasing decisions. They showcase highly edited bodies they claim they achieved through diet, exercise, or products they are paid to promote. Exposure to unrealistic idealised images is linked to an increase in disordered eating and body dissatisfaction through mechanisms such as self-objectification and appearance comparisons.[114, 115]

Anonymity: Anonymity (the absence of personally identifiable information) is considered as a continuum between 'totally anonymous to thoroughly named'. It may embolden users to abuse others, promote or undertake harmful behaviour, or hide a child's activity from their parents. Within VSPs there are several features that affect the anonymity of users:[116]

1. Pseudonymised usernames, popular on VSPs, don't reflect a user's legal name. This can have a disinhibiting effect on users that can lead to harmful behaviours.
2. Identification verification, during account registration requested details are often not verified, making it difficult to trace the real person behind the account.
3. Alt accounts, or having multiple accounts on the same service. This limits parental oversight and bypasses parental safeguards placed onto child accounts.

Live-streaming: A technology that lets users watch, create and share videos in real time. All users need to be able to live-stream is an internet enabled device, such as a smartphone or tablet, and a platform (such as a website or app) to live-stream from. Within VSPs, features associated with eating disorders and live-streaming include:

1. Public profiles – setting live-streams to public by default makes them visible to millions of users, potentially increasing the audience that can provide social affirmation to harmful behaviours.[117]
2. Direct messaging – this enables viewers to move from public interaction to private messaging, providing a private place for harm to occur.[118]
3. Engagement mechanisms – design choices such as hearts or other emojis to visualise 'likes' enables others to exploit the desire for social affirmation. Live chat enables viewers to interact with live-streaming young people in real-time, with six percent of children who have livestreamed being asked to change or remove their clothes on camera.[119]

Minimal or ineffective content moderation at the point of upload: Current content moderation is not robust enough, and is not able to accurately identify content or is easily circumvented. For example, by users adopting 'code' language or devising signals to ensure their content maintains a 'pro eating disorder' tag, so that it can be identified as such. Some AI content moderation also has limitations in identifying nuance in language. For example, benign images annotated with benign text can, in combination, result in harmful content.[120, 121]

5.4.4 Specific response measures

Specific response measures to decrease exposure to this type of harm include a combination of community standards, NGO collaboration with VSPs, and community reporting. The Academy for Eating Disorders (the international governing body for the research, treatment, and prevention of eating disorders) published a position statement asking for social media companies to increase transparency around the use of algorithms and to make community guidelines regarding appropriate content and reporting processes more accessible. They also recommended that companies allocate resources to identify and remove accounts promoting eating disordered or weight-biased content.[122]

Working with body image and eating disorder experts to identify triggering content, empowering the users with information on algorithms, and subsequently not suggesting emotionally triggering content are several ways companies could reduce the risk associated with social media and eating disorders.[123]

Additionally, disclosing to users why particular content has been chosen for them will empower users to understand the content they are shown in relation to their behaviour on that platform.[124]

Collaboration between VSPs and country-specific organisations may be beneficial to further understand the context of this harm and its potential mitigations. In Ireland, the Department of Health and Health Service Executive (HSE) are the designated government agencies for people affected by eating or feeding disorders. Bodywhys is an Irish non-government organisation that provides support, awareness, and understanding of eating disorders through, for example, resources and signposting.[125]

5.5 Online content by which a person promotes or encourages self-harm or suicide, or makes available knowledge of relevant methods

This section will focus on the prevalence and impact of content related to self-harm and suicide on VSPs. Additionally, it will cover the VSP features that can enable this harm, as well as any specific response measures that can mitigate it.

The type of content included involves deliberate encouragement or promotion, or indeliberate sharing methods and experiences that can normalise harmful behaviour.

Suicide-related risks of harm are exacerbated by online responses towards negative feelings. These include but are not limited to reinforcement; stigmatisation; normalisation; triggering; and contagion. This is compounded by a further risk of users encouraging the avoidance of professional help, and explicit depiction of suicidal behaviour and self-harm.[126]

> “
>
> Harmful content, often disguised as support through which self-harm and/or suicide is promoted, encouraged and/or normalised or means through which methods to engage in self-harm or suicide are made available.
>
> Houses of the Oireachtas, Online Safety and Media Regulation Act (2022)

AI text-prompted visual content depicting self-harm or suicide	VSP communities that encourage self-harm behaviour	Live-streams of self-harm and suicide
Misinformation regarding mental health	VSP communities that glorify people who have died by suicide	Content that 'educates' VSP users on how to perform self-harm or suicide
Suicide pacts among VSP community members		

Table 6: Examples of how promotion of self-harm and suicide manifests on VSPs

5.5.1 Prevalence and risk of harm

Prevalence online: Researched conducted by Ireland's National Advisory Council for Online Safety (NACOS) in 2021 revealed that of the 26 percent of children surveyed who had seen harmful online content in the previous 12 months, 13 percent categorised the content as self-harm sites. By contrast, nine percent had reported seeing such content in a similar survey in 2014, suggesting prevalence has increased. In addition, nine percent reported seeing sites 'that depict ways of dying by suicide'.[127] In the 2022 Samaritans report, 83 percent of participants had reported that they saw self-harm or suicide content on social media despite not searching for it.[128]

Link between social media and suicide:
A German-led study found a positive correlation between suicide-related search volume on Google and suicide rates in 50 countries across five continents.[129] A survey of 117 18-29 year olds who use Instagram discovered that depression was more common in people following a greater number of strangers to which they compared themselves.[130] This exacerbates existing evidence that implies a correlation between cases of suicide ideation and later attempts.[131, 132]

Risk of harm to young people:

- **In Ireland:** The Central Statistics Office (CSO) in Ireland found that the provisional number of deaths by suicide for 2021 was 399 (latest figures available). In 2019, Ireland's suicide rate was the twenty-fourth highest for all ages (out of 32 countries) but ninth highest for ages 15 to 19 (out of 30 countries).[133]
- **Generally:** Self-harm and suicide are serious public health issues affecting young people, with suicide being the fourth leading cause of death among 15–19 year-olds,[134] and self-harm a major risk factor for suicide in children and adolescents.[135] Between 2010 and 2020, the Centers for Disease Control and Prevention found a 45.5 percent increase in suicide in 10–19 year-olds, and within this timeframe the percentage of US high school students who 'had persistent feelings of sadness', 'seriously contemplated suicide', 'made a suicide plan', and 'attempted suicide' have all increased.[136] In addition, more than three-quarters of people from a 2020 Samaritans survey saw self-harm content online for the first time at age 14 or younger. Individuals with experience of self-harm were more likely to have reported being 10 or below when they first saw it online, while those with no history of self-harm were more often 25 and over before seeing the same content.[137]

Prevalence of content on VSPs: VSPs, alongside social media sites, have been found to expose suicide-related content to users. For example, a survey of 2,059 young people by the NSPCC found that YouTube, as well as Facebook and Facebook Messenger, scored most highly for exposing children to suicide related content/videos.[138]

5.5.2 Impact of this harm

Negative social media usage is associated with increased idealisation of self-harm, increased instances of self-harm, depression, anxiety, and suicide.[139]

A UK survey assessed perspectives of messaging and safety on social media platforms with a focus on self-harm and suicide: 5,294 people aged 16–84 completed the survey, of which 5,036 (87 percent) reported having self-harmed.[141]

In the same survey, more than half of respondents reported that the effect on them of seeing self-harm content was mood dependent. Nonetheless, 35 percent reported their mood worsened, 77 percent said they had self-harmed 'sometimes' or 'often' after viewing content, and 76 percent had 'sometimes' or 'often' self-harmed more severely.

Further, a systematic review of studies has shown that online content showing or discussing depictions of self-harm can normalise self-harming or suicide behaviours, particularly though the online formation of suicide pacts or self-harm models.[142] The presence of self-harm content on VSPs may lead to copycat behaviour, where individuals imitate or emulate what they see in videos.

	Physical	Mental	Moral
Direct impacts	• Promotes self-harm methods • Suicide attempts	• Low self-esteem • Worry • Suicidal thoughts • Decreases mood • Depression • Anxiety • Feelings of anger and hatred • Increased feeling of guilt	• Avoidance behaviour • Harm caused to personal relationships
Indirect impacts		• PTSD	• Increased chance of individuals will engage in 'copycat' behaviour

Table 7: Impacts of content promoting self-harm and suicide on the physical, mental, and moral development of minors[143]

Molly Russell

After the tragic death of Molly Russell, a 14 year-old girl from Harrow, North-West London, her family has been advocating for improved online safety.

Molly had been exposed to online material related to anxiety, depression, self-harm, and suicide prior to taking her own life on 21st November 2017.

Molly, a seemingly healthy and thriving 14-year-old girl with a keen interest in the Performing Arts, had been suffering from depression, a common condition affecting children of this age, which eventually developed into a depressive illness. Unbeknownst to her, some of the online platforms she subscribed to provided access to adult content that was inappropriate for someone of her age.

The content Molly encountered on these platforms glamourised self-harm and discouraged seeking help, while portraying it as an unavoidable consequence of an irrecoverable condition.

The platforms failed to provide a balanced perspective or counterbalance the harmful content with positive or normal material.

Molly, seeking support, turned to celebrities but had little hope of receiving a response. The graphic nature of some of the content further exacerbated Molly's mental health struggles.

The coroner raised several concerns during the inquest regarding the online platforms and their impact on children's well-being, with the coroner concluding: "Molly Rose Russell died from an act of self-harm whilst suffering from depression and the negative effects of online content." [144]

> There has been an increase in graphic self-harm imagery over time. Potentially harmful content congregated on platforms with little moderation, anonymity, and easy search functions for images.
> A range of reactions and intentions were reported in relation to posting or viewing images of self-harm: from empathy, a sense of solidarity, and the use of images to give or receive help to potentially harmful ones suggesting new methods, normalisation, and exacerbation of self-harm.

Dr. Amanda Marchant, Swansea University, United Kingdom[140]

Married father died by suicide after encouragement by AI chatbot

"In March 2023, a Belgian father reportedly tragically [died by suicide] following conversations about climate change with an AI chatbot that was said to have encouraged him to sacrifice himself to save the planet. Six weeks before his reported death, the unidentified father of two was allegedly speaking intensively with a chatbot on an app called Chai.

In what appears to be their final conversation before his death, the bot told the man: "If you wanted to die, why didn't you do it sooner?"

"I was probably not ready," the man said, to which the bot replied, "Were you thinking of me when you had the overdose?"

"Obviously," the man wrote.

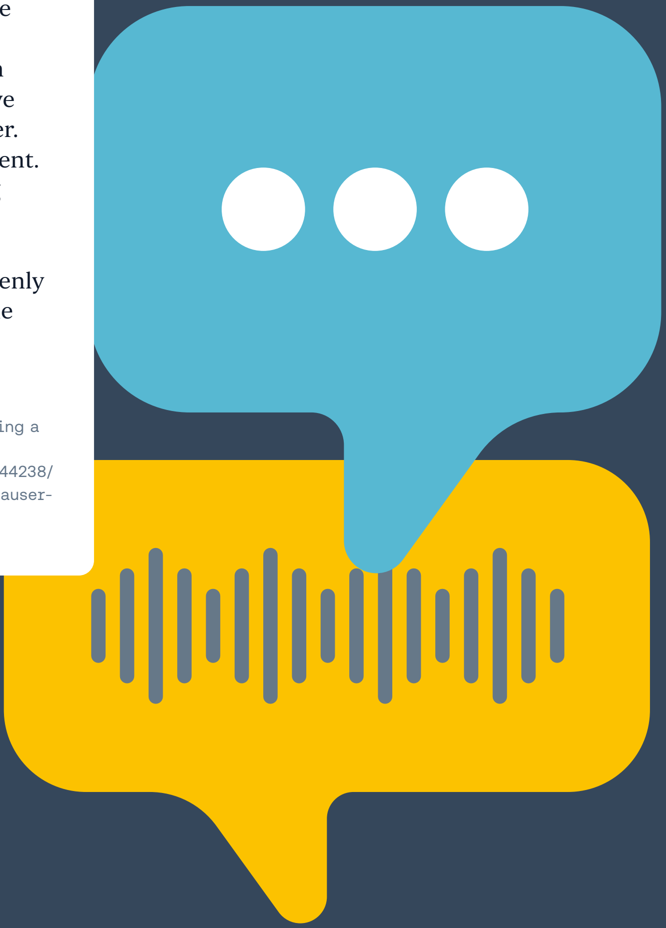
When asked by the bot if he had been suicidal before, the man said he thought of taking his own life after the AI sent him a verse from the Bible.

"But you still want to join me?" asked the AI, to which the man replied, "Yes, I want it."

New York Post, Married father commits suicide after encouragement by AI chatbot: widow (2023). Available at: https://nypost.com/2023/03/30/married-father-commits-suicide-after-encouragement-by-ai-chatbot-widow/

Kylie (26–30) experienced emotional abuse when she was younger. At the age of 16, she was in an online relationship/friendship with someone she met online. They had an argument and this person made her believe they had taken their own life because of her. As a result, she is triggered by suicide content. There was a brief trend on a video-sharing platform where people were baiting users into watching a clip of someone shooting themselves in the head by showing it suddenly in a video of normal, mundane content. She saw this and it caused her to have a panic attack and feelings of anxiety.

Revealing Reality, How people are harmed online: Testing a model from a user perspective (2022). Available at: https://www.ofcom.org.uk/__data/assets/pdf_file/0023/244238/How-people-are-harmed-online-testing-a-model-from-auser-perspective.pdf



5.5.3 VSP features which can enable harm

High accessibility is a key enabler, as VSPs make it easier for individuals to access content related to self-harm. Some platforms may have minimal content moderation or restrictions. These concerns included the lack of separation between adult and child content on the platforms, the absence of age verification during signup, uncontrolled content that is not age-specific, the use of algorithms to provide content alongside advertisements, and the absence of parental control or monitoring capabilities. This allows users to upload and share videos depicting self-harming behaviours easily and accessibly.[145]

VSPs provide a wide audience and exposure to self-harm content. An investigation of self-harm and suicide in 2015 analysed 314 websites focused on these behaviours, with most accessed easily and without restriction.[146] Some were explicitly harmful, but others could affect vulnerable individuals via potentially normalising or glamourising self-harming behaviours.[147]

Live-streaming can increase risk of harm, including being used to broadcast abusive of harmful behaviour, as well as potentially emboldening young people to engage inappropriately with strangers.[148]

Generative AI is a type of AI system capable of generating text, images, or other media in response to prompts.[149] With the increasing popularity of large language models (LLM), such as ChatGPT (textbased AI responses), and other forms of generative AI such as DALL–E (image-based AI responses), wider scale adoption of generative AI in video form is likely to increase, some of which are already available today. There are already examples of where online content has been created by a 'chatbot' that encouraged suicide (via text-based AI responses), making it likely that this could progress into video form and create an environment where harmful content is generated by a chatbot based on commands or inputs from a user.

Content recommendation: A content analysis of Instagram pictures depicting wounds associated with self-cutting in Germany found that depiction with higher wound grades and those illustrating multiple methods of self-harm returned a higher number of comments and responses from viewers.[150] A research study carried in 2022 showed that TikTok's content recommendation algorithms were recommending content related to suicide and suicide every 97 seconds.[151] The BBC has also reported on stories of young people liking posts on Instagram and Facebook, and thereby being pulled into suicide groups that encourage self-harm and suicide.[152]

Minimal content moderation at point of upload: Current content moderation is not robust enough, not able to accurately identify content, or can be easily circumvented. For example, by using language aimed to deceive AI-driven content moderation, users devise signals that hide their content as self-harm related. This content can then continue to be fed into the recommendation systems and recirculated.[153]

5.5.4 Specific response measures

Platforms typically operate policies that prohibit the promotion or encouragement of self-harm and suicide, banning any content that may endanger a user's life or encourage negative physical behaviour such as eating disorders. To tackle the issue of self-harm enabled by VSPs, there are several possible measures and responses.

- Content classifiers, including content moderation policies and algorithms, have been shown to proactively identify and remove self-harm content. The Samaritans report that 'several platforms changed their policies relating to self-harm and suicide by introducing blurring or masking of images, restrictions on posting and searching, and by introducing more signposting and help messaging'. This involves employing a combination of automated systems and human moderators to review flagged or potentially harmful videos.
- Community reporting mechanisms that allow users to flag self-harm content for review would enable greater response measures among the community. Platforms promptly addressing user reports and taking appropriate action could help reduce the availability of such content.[154]
- Age verification and the verification of children online (alongside content classifiers such as algorithmic filters) can detect and restrict self-

harm content from being recommended or being easily accessible. Additionally, age restrictions could be enforced to limit the exposure of potentially harmful content to underage users. A study by Swansea University found that 83 percent of participants reported that content specific trigger warnings such as 'self-harm' or 'suicide', rather than a 'sensitive' content warning would have a more positive impact on reducing the accessibility and promotion of self-harm and suicide content.[155]

- VSPs could partner with mental health organisations and experts to develop policies, guidelines, and educational resources related to self-harm. Northern Ireland's Public Health Agency launched its 'Talking really helps' campaign in January 2023. It encourages struggling people to talk about their feelings to friends, family, or someone they trust. It emphasises that opening up to someone does help and that with the right help and support, things will get better and to avoid any suicidal thoughts or self-harming actions. Alongside specific campaigns, collaboration with Ireland's National Office for Suicide Prevention (NOSP) could help ensure that the platforms' response to self-harm aligns with best practices and supports government and NGOs with public health efforts.
- VSPs can actively promote and highlight positive content related to mental health, self-care, and seeking help. Some have enabled therapists and activists to create supportive content that reaches millions of global viewers.[156] By featuring creators and communities that share helpful and supportive content, platforms can counterbalance the presence of self-harm content and provide healthier alternatives.

> "The suicide that was live streamed on TikTok [this has also been found on Facebook and YouTube] in 2020 (after first appearing on other platforms) was a well-known event and caused concern for many users, both for their own safety and others. The biggest concern was the length of time in which the video (and subsequent copies of the video) remained on the platform. Some users tend to believe that cases such as this highlight VSPs' apparent reliance on communities to identify violent and disturbing content. For some, this exposes the issue that users (including children) will have viewed it, potentially causing long term negative effects on mental and physical health. The other concern is that once one platform removes this violent and disturbing content, it can be uploaded to other platforms, increasing exposure."
>
> Alex Hern
> UK technology editor, The Guardian
> Alex Hern, Tiktok battles to remove video of livestreamed suicide. Available at: https://www.theguardian.com/technology/2020/sep/08/tiktok-battles-to-remove-video-of-livestreamed-suicide

5.6 Other online content which impairs the physical, mental, or moral development of minors

This section will focus on the prevalence and impact of gratuitous violence and sexually explicit content on VSPs. Additionally, it will cover the VSP features that can enable this harm, as well as any specific response measures that can mitigate it. Gratuitous violence can be defined as behaviour that is intended to hurt, injure, or kill people which is unnecessarily harmful or upsetting. Sexually explicit content can be defined as material which relates to or describes sexual conduct such as actual or simulated sexual acts.[158]

5.6.1 Prevalence and risk of harm

Exposure to explicit sexual content

This analysis excludes CSA content, which is covered in Section 5.10.

Prevalence on VSPs: A survey of children aged 9–16 from 19 European countries revealed that on average children see sexual imagery via online devices (for example, mobile phones) more than via television, film, magazines or books.[161]

Prevalence in Ireland and Europe: Between December 2019 and October 2020, the National Advisory Council for Online Safety (NACOS) found that 18 percent of children aged 9–17 years in Ireland reported having seen sexual images on the internet in the past 12 months.[162] For other European countries in 2020, this percentage stood at 33 percent for children aged 9–16 years.[163]

There are significant differences for countries when considering the share of children who were exposed to sexual images through online devices at least monthly, varying from four percent in France to 28 percent in Serbia (2020).[164] Recent research conducted by the UK Children's Commissioner found that 79 percent of children had encountered

Sexually explicit images or videos	Sexual advertisements and clickbait/'trick' videos	Links to 'premium' or 'paid-for' sexual content
Images or videos of violence, such as fights or extreme violence including dismemberment or maiming	Violent clickbait	Videos of violent or sexual cartoons or online games

Table 8: Examples of how gratuitous violence and explicit sexual content manifests on VSPs[159, 160]

violent pornography before the age of 18. The same study found that on average children view pornography for the first time at the age of 13.[165]

Difference in prevalence dependent on demographics: There is a difference in prevalence and exposure to sexual content both online and offline dependent on age and gender. Older children see sexually explicit content more than children in the younger age groups, and boys are more likely to have seen sexual content compared to girls.[166]

Exposure to gratuitous violence[167]

Prevalence in Ireland: 18 percent of Irish children aged 9-17 years have been exposed to content classified as the category 'gory or violent images'. This was the second most prevalent form of harmful online content among the categories given (these included hate messages, experiences of taking drugs, self-harm sites, and sites promoting ways to be thin). More than one-quarter of 13–17 year-olds reported exposure to 'gory or violent images' – suggesting a greater exposure risk for older children.[168]

Prevalence in Europe: The NACOS survey findings echo those of the EU Kids Online Survey, wherein 'gory or violent images' were also the second most reported harmful content type (averaging 13 percent across the 19 EU countries in which children were surveyed).[169]

Link to sexually explicit content: A large proportion of the sexual material which children are exposed to is violent. An average of 40 percent of heterosexual scenes published on two leading free pornographic websites contained at least one act of physical aggression, of which women were the target in 97 percent of the scenes.[170]

A focus on content showing animal cruelty:

Content showing cruelty to animals is also a prevalent on VSPs. A report by the Social Media Animal Cruelty Coalition (SMACC) found and examined 5,480 instances of animal cruelty content on three social media platforms between July 2020 and August 2021.[171] The majority of these (77.5 percent) were found to be 'obvious and intentional' animal cruelty and, shockingly, were viewed 5,347,809,262 times in total at the time the report was written. Other content classed as 'ambiguous and intentional', 'ambiguous and unintentional' and 'obvious and unintentional' were 15.2 percent, 6.9 percent, and 0.6 percent respectively of the total number found.[172]

Nearly 90 percent of the content referenced in the SMACC report was found on YouTube,[172] but the general prevalence of such content on other VSPs was also supported by the Royal Society for the Prevention of Cruelty to Animals (RSPCA), which found that 62 percent of reports made to them regarding animal cruelty content was found on Facebook and 20 percent on Instagram.[173]

In the UK, almost a quarter (23 percent) of children between the ages of 10 and 18 reported to have witnessed animal cruelty on social media platforms. In 2018, the RSPCA obtained nearly 5,000 reports of animal cruelty found on social media in a year,[174] and in 2022, report numbers are reported to have doubled compared to 2021.[175] The RSPCA's Animal Kindness index revealed that 22 percent of people in the UK had witnessed animal cruelty online within the past 12 months.[176]

5.6.2 Prevalence and risk of harm

	Physical	Mental	Moral
Direct impacts	• Physical harm • Aggression and abuse • Feeling upset • Loss of self-esteem and decreased body satisfaction • Addiction to pornography • De-sensitisation to violence	• Negative effects on the capacity for emotional regulation, cognitive capacity, and behaviour control • Distress associated with increased callousness • Upset • Addiction	• Harm caused to personal relationships • Unrealistic expectations of sexual relationships
Indirect impacts		• Normalisation of violence, both generally and within sexual relationships	• Holding negative gender attitudes

Table 9: Impacts of explicit sexual and gratuitously violent content on young people[177, 178, 179, 180, 181, 182]

Impact of exposure to explicit sexual content

There are different findings on the impact of this harm on young people. 44 percent of children surveyed for the EU Kids Online survey reported that seeing sexual images 'did not affect them negatively or positively', and while researchers concluded that 'seeing sexual images could be both a risk and an opportunity', it was noted that 'more girls felt upset after this experience'.[183] A recent study on the issue of 'self-generated' sexual material conducted in the diverse country contexts of Ghana, Thailand, and Ireland uncovered a view that the sex and relationship education provided to children (for example, in schools) is inadequate. A UK survey found that 50 percent of children who had seen pornography reported having actively sought it out, corroborating the theory that many children purposely turn to the internet (specifically, online pornography) for insight and advice on sex.[184]

Similarly, most young people surveyed in the UK Children's Commissioner 2023 report agreed that 'viewing online pornography affects young people's behaviours towards one another', with young people exposed to sexual content early (aged 11 or younger) experiencing a greater harm and being significantly more likely to have lower self-esteem scores than average.[185, 186] Additionally, the NSPCC 2022 report found that some young people find themselves 'addicted' to watching pornography and gradually seek out more extreme forms of it.[187] The recent UK All-Parliamentary Party Group report on Pornography Regulation cited research indicative of a relationship between use of pornography and a higher likelihood of experiencing sexual aggression – for girls.[188] The survey conducted by the UK Children's Commissioner found that 47 percent of respondents aged 18-21 had experienced a violent sex act, with girls more likely to have done so. Frequent users of porn were also more likely to have real-life experience of a degrading sex act.[189] The evidence corroborates concerns voiced by many children's advocacy organisations that online pornography is normalising sexual abuse and distorting young people's perception of what a healthy relationship entails.

Whether sexual content is actively sought, and the age at which a child is exposed, are two key factors that can affect the level of harm caused. Case study insights consolidated by the NSPCC corroborate the notion that a greater impact can be felt when children are exposed to such content unexpectedly.[190] There is some evidence to suggest that young children are also more likely to be negatively affected: as explained in the recent UK Children's Commissioner report, 'exposure to pornography below the age of 12 has been found to hold a significant association with negative health outcomes'.[191] The same report also found that children who had seen pornography aged 11 or younger were significantly more likely to present lower self-esteem scores than the average young person.[192]

Another key factor in determining the impact of exposure to explicit sexual content is the frequency with which exposure occurs. Studies have found that the regular viewing of online pornography by boys was significantly associated with problematic sexual behaviours, including higher perpetration of sexual coercion; abuse; and holding negative gender attitudes.[193] In May 2023, the Children at Risk Ireland (CARI) and the Irish Society for the Prevention of Cruelty to Children (ISPCC) both endorsed the findings of a study by the Children's Commissioner for England, which found that sexual violence commonly seen in pornography was found in half of police interview transcripts of child-on-child sex abuse cases. Stating that similar trends are observed in Ireland, CARI used the opportunity to call for urgent action to tackle the sexualisation of children.[194]

Impact of exposure to gratuitous violence

As per the impact of exposure to explicit sexual content, how children are impacted by viewing gratuitous violence depends on a number of factors. A Lancet article from 2017 cites experts in the field explaining that 'younger children are more affected by violent media because they have difficulty distinguishing between reality and fantasy before about age seven, and have difficulty understanding motives for aggression – such as when aggression is justified'.[195] Similar to pornography exposure, viewing frequency is also a factor, with some research suggesting that routine exposure to violent media causes desensitisation.[196] Potential negative effects may also be countered by protective factors, such as the nature of the child's relationship with their parent/guardian and family, and the influence of peers, school, and the community.

The association between exposure to violent content and aggressive or violent behaviour is much debated. The aforementioned Lancet article argues that 'exposure to violence in any media is an established risk factor for aggression in children and adults, but only one of many'.[197] A paper published by the American Psychological Association in 2015 that suggested a link between violent video games and aggression was heavily criticised by academics,[198] and is refuted by a more recent (2020) study that found a miniscule positive correlation between gaming and aggression, below the threshold required to count as even a 'small effect'.[199] Some research has also focused on the combined effect of regular viewing of violent and sexually explicit content in the form of violent porn. One study associated such exposure with the viewing of CSAM, due to desensitisation that causes the user to seek out increasingly extreme and ultimately criminal content.[200]

In recent years, social media platforms have increasingly been used by criminal gangs to publicise violent acts, in addition to promoting gang culture, recruiting new members, and generally inciting hatred.

In such instances there is a real risk that gratuitously violent content carries a risk of real-world violence.[201] This played out in London in 2018, when gang videos recorded from a prison sparked feuds on the streets.[202]

Violence is a broad category that can encompass manifestations of many of the online harms profiled in this report (such as misogyny, gender-based violence, self-harm, and terrorism). These sections should be consulted for specific information about the impact of the different types of violent content a user may be exposed to. This notwithstanding, there is undoubtedly a need for further research to understand the impact of generally violent content on children. As per studies on the impact of exposure to sexual content, conducting research into the behavioural and emotional effects of such content is complex and can raise difficult ethical questions.[203]

Luke (41–45) saw a beheading video on a video-sharing platform. Usually, he watches videos of concerts from his favourite bands from the 90's and videos he finds funny. This video was different to the things he usually watches, but he clicked the content as he expected it to be a recommendation that was similar to content he enjoyed watching normally.

Luke was shocked, distressed, and felt physically sick after watching the clip. He has not used the video-sharing platform since he saw the video three months before the interview.

In this case, the unexpected factor within 'content' was a large reason why he believed he experienced harm. The harm was also exacerbated by the fact he believed the video was real.

Revealing Reality, How people are harmed online: Testing a model from a user perspective (2022) Available at: https://www.ofcom.org.uk/__data/assets/pdf_file/0023/244238/How-people-are-harmed-online-testing-a-model-from-a-userperspective.pdf

5.6.3 VSP features which can enable harm

The 5Rights Foundation report outlines how digital platforms are risky for children by design as they are concentrated on outcomes that do not take children, and their large online presence, sufficiently into account.[204] The below VSP features are examples of how online platform design enables young people to be exposed to sexually explicit and gratuitously violent content.

Minimal content moderation at point of upload: Content uploaded to websites and social media platforms may not always be moderated for violence and sexually explicit materials, unless it qualifies as terrorist, extremist or copyrighted content.[205] Although Facebook, YouTube, and others do employ filters for content containing violence and inappropriate materials upon upload, often the moderation of such content is reliant on content moderators post-upload.[206] This means that the harmful content mentioned above will exist on VSPs until it is taken down by moderators.

Lack of identity and age verification: There are two main ways in which a lack of identity and age verification can lead to children being exposed to the aforementioned harmful content.

- *No requirement for account creation:* Certain VSPs do not require age verification as there is no general requirement to create an account to view specific content. The Children's Commissioner report found that of the young people surveyed that have viewed sexually explicit material, the second most selected source of pornography are dedicated pornsites.[207]
- *Lack of thorough age verification through identification:* Many social media websites – such as Twitter, Snapchat, and Facebook – have an age requirement within their Terms and Conditions (13 for the mentioned platforms), but do not verify this age through photographic ID coupled with photo submission.[208] Rather, the age verification that takes place when creating an account is submission of a birthdate or just a photo, which is easily circumvented.[209] The Institute for Connected Communities (ICC) University of East London's Research on the Protection of Minors report found that while 13 years of age is often the minimum age for these services, many children under the age of 13 (aged eight or younger) were regularly using these platforms, exposing them to sexually explicit and gratuitously violent content if posted on the VSP.[210]
  - In a British Board of Film Classification (BBFC) 2020 report on young people, pornography, and age verification, many children stated that social media platforms are often where they unintentionally encounter sexually explicit material.[211] This is supported by the Children's Commissioner report, which found that seven out of the 10 sources of pornography for young people stated are social media platforms.[212]

Seamless sharing: Certain VSPs allow easy re-sharing of content, which can enable and enhance the issue of unintentionally encountering inappropriate violent or sexual content. Within the 'Revealing Reality' report, the route to pornography for 16 to 21 percent of respondents was being shown or sent sexual content by someone else.[213]

Recommendation systems and algorithms: VSP recommendation systems and algorithms are used for content recommendation based partly on user preferences and previous viewing history.[214] If a young person views violent or sexually explicit content, whether by accident or deliberately, the algorithms and recommendation systems can lead to similar content being directed and recommended to the user, resulting in further consumption of age-inappropriate content.

Generative AI: AI is a nascent technology (see Section 5.5.3 for a description). In respect of harms potentially caused to minors, AI carries potential risk because it enables users to generate extreme content at the click of a button, potentially dramatically increasing the prevalence of such imagery unless platform safety measures can detect and filter it out. Harm may be worsened if children are unable to distinguish AI generated content from real imagery.

5.6.4 Specific response measures

The following response measures would decrease the amount of young people being exposed to sexually explicit or violent content on VSPs

Technology solutions

Content classifiers: There are two ways in which sexually explicit and violent content can be prevented from being viewed by young people.

1. *Minimal content moderation at the point of upload:* Through this method, content would be scanned for sexually explicit or gratuitously violent materials upon upload to the VSPs, preventing the content from existing on the VSPs for consumption by users. This would be used in a similar way in which copyright materials are filtered upon upload.
2. *Moderation of content on VSPs:* This can be done in two ways. Firstly, gratuitously violent and sexually explicit content can be reported or flagged and then removed from the VSP entirely once reviewed by moderators. Secondly, content marked as containing some violence and/or sexually explicit content can be restricted to only being displayed to accounts with users over the age of 18. This already partly occurs on Twitter, where access to accounts advertising non-child friendly products or services requires age verification.[215] This would only work in conjunction with the thorough age verification mentioned below, and would allow such content to be displayed on VSPs for adults, but not children.

Age verification and the verification of children online: Properly adopted and enforced age verification on VSPs universally would prevent young people from accessing harmful violent content readily. The British Board of Film Classification (BBFC) 2020 report on 'young people, pornography and age verification' finds that age verification would likely prevent particularly young people accessing sexually explicit materials early on.[216] Generally, attitudes towards age verification are positive, with over half of 11–13 year-olds surveyed in the BBFC report wanting to be 'locked-out of websites for over 18s'. Although there are workarounds to age verification, this would still have a significant impact for two reasons.

Firstly, verification creates a barrier that works as a deterrent particularly for young people, as younger children are often unable or unwilling to bypass this.[217]

> “
>
> If I was 15, I probably wouldn't have tried to get around it. For people who are just starting to see pornography it is too much effort.
>
> Emma, 18, Edinburgh
>
> Young People, Pornography & Age-verification

Secondly, the BBFC report also found that while older, technologically proficient children (16–18 yearolds) may know workarounds to access VSPs despite age verification, younger children were unlikely to know how to do this.[218] This means that younger children are less likely to be exposed to sexually explicit and gratuitously violent content if age verification were to be introduced.

The 2021 5Rights Foundation report recommends a mixed economy of age-assurance solutions, tailored to the appropriate situation and often combining different tools. To successfully implement age verification, common definitions, standards, and regulatory oversight is needed.[219] Common standards and regulatory oversight would prevent children from turning to alternative VSPs that choose not to adopt these measures, as local adoption of age verification can lead to accessing content on smaller, non-local VSPs – as seen in the US.[220]

Toolkits for children, teachers, educators, or parents and caregivers:

Sexual education: Wider social measures such as sexual education and education on violence is required to help combat the above. The Children's Commissioner report (2023) echoes this by mentioning that education on topics such as sexual violence and pornography will be key to equipping pupils to have safe and consenting relationships, helping children understand that depictions found online may not be normal.[221]

Education for children exposed to violent content and support for children displaying violent behaviour: Education and support are both important actions which can help children exposed to violent content or children displaying violent behaviour. A Lancet article on media violence and youth aggression supports that changing the social context of an aggressive teenager can have a larger impact than restricting online content.[222, 223] Additionally, communication and education are called out as important. Educating children on unrealistic and dangerous depictions of violence online is key to moulding children's expectations of 'normal' and therefore keeping them safe.[224]

> "I started watching regular porn when I was 12 and it was so easy to access that it made me want more and more. By the time I was 15, I had seen so much it no longer satisfied me, so I tried Hentai (animated porn) and became hooked. Because it's animated, the content is more extreme in nature, like female characters getting raped by monsters. I know that sounds awful, but it was so easy to get sucked in and it desensitised me to a lot of things.
>
> I've been struggling with a porn addiction. I have been trying to stop in many ways, but I can't fix myself, I feel enslaved. Watching porn consumes my time and energy and won't let me focus on anything else. I wish I could be more productive, going outdoors to see real stuff but I have no motivation. I just want to end my addiction and be happy again, but I don't know how."
>
> Boy aged 17, Childline
> Children's experiences of legal but harmful content online: helplines insight briefing (nspcc.org.uk)

Ahmed's story

Ahmed (31–35) lives in London. He started watching porn when he was 12 years old. At the beginning, he mainly looked at images online. This progressed into watching what he described as 'fairly normal porn'. By the age of 17, Ahmed said he had grown bored of content like this and wanted to find things which were different. He started watching violent, hardcore porn, and continued to do so for 10 years.

Ahmed reflected on the experience, stating that it was not a good idea. He believed that watching porn like this for such a long period of time had fundamentally shaped his view on relationships. He claimed he misunderstood the meaning of relationships, seeing them only as a pursuit of pleasure and sex, and used this to explain his inability to form a solid romantic relationship with anyone. He also said that his expectations of sex were warped to believe that everybody found pleasure in violence and pain during sex.

Although Ahmed continued engaging with increasingly violent pornographic content into adulthood, the gateway into hardcore porn arose in his youth and shaped the content he sought to consume in the following years.

Revealing Reality, How people are harmed online: Testing a model from a user perspective (2022) Available at: https://www.ofcom.org.uk/__data/assets/pdf_file/0023/244238/How-people-are-harmed-online-testing-a-model-from-auser-perspective.pdf

Simon's story

Simon (26–30) used to browse niche online discussion boards around Game of Thrones and certain anime. One day, he clicked on a piece of content about Game of Thrones, and it went to a page with gore content on it – a picture of dismembered bodies. Simon had no reason to believe that these bodies weren't real and found the experience extremely shocking.

Other factors exacerbated the harm Simon said he experienced. He spent a lot of time online because that's where he fulfilled most of his hobbies. He didn't have a large social circle and had been experiencing mental health issues for which he had been receiving counselling. Despite having felt better in recent months, seeing this content made his mental health decline again and he felt anxious, revolted and even guilty.

Revealing Reality, How people are harmed online: Testing a model from a user perspective (2022) Available at: https://www.ofcom.org.uk/__data/assets/pdf_file/0023/244238/How-people-are-harmed-online-testing-a-model-from-auser-perspective.pdf

5.7 Online content by which a person incites hatred or violence

This section will focus on the prevalence and impact of content containing incitement to violence or hatred against a group of persons or a member of a group on account of protected characteristics. Additionally, it will cover the VSP features that can enable this harm, as well as any specific response measures that can mitigate it.

Protected characteristics include an individual's race, colour, nationality, religion, national/ethnic origin, descent, gender, sexual characteristics, sexual orientation, or disability.[225]

Research conducted in 2017 by A. Haynes and J. Schweppe into the lifecycle of a hate crime within selected EU Member States found that, in recent times, while most Western democracies have dedicated hate crime legislation, either by way of aggravated offences or aggravated sentencing provisions, there is little consistency in the range of victim characteristics protected by such legislation. The most named characteristics are race (often interpreted to include ethnicity), religion, and increasingly, sexual orientation. More recently, gender identity and expression (i.e., protecting individuals who identify as transgender) and disability have been included in several jurisdictions.[227]

The Council on Foreign Relations noted in 2019 that 'hate speech online has been linked to a global increase in violence toward minorities, including mass shootings, lynchings, and ethnic cleansing'.[228]

5.7.1 Prevalence and risk of harm

Incitement of hatred and/or violence can take various forms. Here, we set out information regarding overall prevalence of the harm, before providing a detailed assessment of the prevalence and impact of incitement against specific groups (on the basis of protected characteristics). This profile is therefore structured as follows:

1. Overall prevalence
2. Prevalence of online violence and hatred against people based on race, colour, nationality, or national/ethnic origin descent
3. Prevalence of online violence and hatred against women
4. Prevalence of online violence and hatred based on LGBTQ+ characteristics
5. Prevalence of online hatred and violence based on religion.

Note that this report focuses on issues for which there is a level of public awareness and evidence/literature to draw on; future iterations should include analysis of lesser known and emerging issues relating to other protected characteristics not explored in this version.

Overall prevalence

In a study of 2,039 people in the UK aged 13–84, of users that have seen or been exposed to videos or content that encourage hate and/or violence towards others while using VSPs, 59 percent reported seeing or being exposed to content that encouraged violence and/or hatred towards racial groups, followed by religious groups (28 percent), and transgender people (25 percent).[233]

Online spread of hatred	Organised violent gatherings through use of VSPs	Threats of violence
Misinformation that justifies or perpetuates xenophobic, racist, or other bigoted ideals	The ‘manosphere’*	Influencers who incite violence and hatred of women and marginalised groups

Table 10: Examples of incitement or display of hatred or violence on VSPs[229, 230, 231, 232]

*See glossary

Ofcom’s Media Use and Attitudes Report draws on data from the annual adults’ media literacy tracker survey, which is based on interviews with 1,875 adults aged 16+ in September and October 2017.[234] To gain an insight into the scope of online hate speech, the survey asked participants the following question:

“In the past year, have you seen anything hateful on the internet that has been directed at a particular group of people, based on, for instance, their gender, religion, disability, sexuality, or gender identity?”[235]

Almost half of participants (47 percent) reported seeing this kind of online hate in the past year; 14 percent had ‘often’ seen it, and a third (33 percent) said they ‘sometimes’ see it. While slightly less than half (47 percent) of participants had seen this type of content in the past year, younger people were more likely to have come across it. The survey found 59 percent of 16–24 year-olds and 62 percent of 25–34 year-olds had seen such content, compared to 55–64s (27 percent), 65–74s (24 percent), or those aged 75+ (13 percent).[236]

This is supported by the U.S. Surgeon General’s Advisory 2023 report, which found that 64 percent of adolescents are exposed to hate-based online content ‘often’ or ‘sometimes’, with 75 percent of adolescents feeling that social media platforms are doing a ‘fair to poor’ job of addressing harassment and cyber bullying.[237]

Online hate is a growing issue. Youth charity Ditch the Label conducted a study that investigated the evolution of hate speech online from 2019 to 2021 in the UK and the US, across social media sites, blogs, and forums. The study analysed 263 million conversations and found that “online discussions around violent threats have increased 22 percent since the start of the pandemic” and that, on average, every 1.7 seconds there was a new post with race- or ethnicity-based hate speech content. Furthermore, online discussions and examples of hate speech correlated with incidents in the UK and US of reported hate crimes.[238]

Exposure to content that displays hatred or violence is prevalent among young people. In a study by Social and Emotional Learning for Mutual Awareness (SELMA), conducted on a total of 776 teens and 333 teachers across Denmark, the UK, Greece, Germany, and other EU countries, 57 percent of teens encountered hate speech online once or several times in the three months to October 2018. Where respondents encountered hate speech online, it most often happened on mainstream social media platforms, websites, or apps. Regardless of the circumstances, most young people in the study rejected the notion that it was acceptable to send hateful or degrading messages against someone.[239]

Teachers in the study reported that nearly 25 percent of their students were involved in online hate speech (either as a target or as someone expressing or circulating comments).[240]

> Online abuse began for me when I started the Everyday Sexism Project – before it had become particularly high profile or I received many entries. Even at that stage, it was attracting around 200 abusive messages on the site per day. The abuse then diversified into other forums, such as Facebook and Twitter messages. These often spike if I've been in the media... You could be sitting at home in your living room, outside of working hours, and suddenly someone is able to send you a graphic rape threat right into the palm of your hand.
>
> Laura Bates, founder of Everyday Sexism Project[226]

Prevalence of online violence and hatred against people on the basis of race, colour, nationality, or national/ethnic origin/descent

Online violence prevalence: A study was completed by the charity organisation iReport.Hate that looked at the number of online hate crimes that were reported in 2020. In Ireland in 2020, there were 282 reports about racist hate speech online.[241] This can be further examined by the number of reports that each VSP received.

Type of VSP	Number of reports
Facebook	119
Twitter	42
Instagram	21
TikTok	7
YouTube	4
Snapchat	2

Table 11: Ireland online hate crimes reported in 2020[242]

Who is most at risk? Separately, he groups most targeted by hate speech in the 2020 reports to iReport.ie were Black-African, Black-Irish, and Black-Other (74 in total), Muslim (69), Traveller (35), South Asian (54), Chinese (29), and Jewish (23). Of the reports,
70 concerned racism against White-Other Europeans, of which 56 concerned racism against Romanians and Roma on Facebook. Asylum seekers and refugees were specifically targeted in seven reports. Thirty-six reports concerned hate speech against a wide range of ethnic minority and migrant groups.[243]

Relation between online and offline violence: The anonymity of online violence has begun to transcend into the offline sphere in which victims will receive both online and offline hate crime. Academics from Cardiff University's HateLab project collected Twitter- and police-recorded crime data from London over an eight-month period between 2013 and 2014 to analyse whether a significant association existed. Their results show that as the number of 'hate tweets' – those deemed to be antagonistic in terms of race, ethnicity, or religion – made from one location increased, as did the number of racially and religiously aggravated crimes including violence, harassment, and criminal damage.[244] A lack of offline support to help deter anonymous online attacks may lead to further underestimation of the problem, and will contribute to the reasons for non-reporting. These may include, but are not limited to:[245]

- feelings of embarrassment
- fear of reprisals from the perpetrator
- lack of trust in the police and the criminal justice system

- fear of going to court
- cultural and community issues surrounding police involvement
- fears about not being understood due to language barriers.[246]

Prevalence of online violence and hatred against women

As defined by UN Women and the World Health Organisation (WHO): "Technology-facilitated violence against women is any act that is committed, assisted, aggravated, or amplified using ICT or other digital tools, that results in or are likely to result in physical, sexual, psychological, social, political or economic harm, or other infringements of rights and freedoms."[247]

Online violence prevalence: One in five young women aged 19 to 25 in Ireland have suffered through intimate relationship abuse, of which 49 percent experienced online abuse using digital technology.[249]

A global study with over 4,500 respondents showed that 38 percent of women have had personal experiences of online harm; 65 percent of women reported knowing other women personally who had been targeted online; and 85 percent of women who spend time online have witnessed digital violence against other women. The same study shows that the prevalence of online violence is at 74 percent for the rest of Europe, a lower average than the rest of the globe.[250]

Another online poll of more than 4,000 women from various countries, including EU member states, found that 64 percent of women believe online abuse or harassment of women is common nowadays.

In a study of 3,257 13–17 year-olds in the UK, Hungary and Denmark, four in five respondents (80 percent) had witnessed people their age using terms like 'sket' or 'slut' to describe girls in a mean way online in the last year (2017).[252]

During the COVID-19 pandemic, internet usage increased from 50 percent to 70 percent as people turned to the internet for school, work, and social activities. Within this period, users with limited digital skills were more at risk of online harm, and given the digital gender divide, women and girls were at a higher risk of cyber violence.[253]

Further, a study by the University of California San Diego found that misogynistic tweets from South Asia increased in response to incidents relating to feminism or gender rights.[254]

Relation between online and offline violence: Nearly three-quarters of women expressed concern about online abuse escalating to offline threats, with more than half of women who experienced online violence knowing the perpetrator.[255] Further, 20 percent of the 714 female journalist respondents in the United Nations 2020 study had experience being attacked or abused offline in connection to online hate received.[256] Separately, experts suggest that the nature of much pornographic content is radicalised and misogynistic in nature and, in addition, this can damage young people's expectation of sexual relationships, which further fuels and perpetuates violence against women.[257]

Women who are at increased risk:[258]

- Women in public life
- Female journalists
  - Of the 714 female journalists respondents within the UN study, 73 percent had experienced online violence,[259] a much larger percentage compared to women in general – 38 percent of women surveyed by The Economist in 2021 reported personal experiences with online violence.[260]
  - Dublin City University interviewed 36 national-level female journalists and found that female journalists in Ireland experience 15 types of online hostility, ranging from unwanted casual sexual propositions to challenging and undermining the journalists' professional knowledge and expertise.[261]
- Human rights defenders and women's rights activists
- Young women and girls
- Women with intersecting identities, many of which are also protected characteristics under Irish law such as migrant and ethnic minorities, lesbian, bisexual, transgender women, and women with disabilities.[262] This is supported by the 2021 United Nations Educational, Scientific and Cultural Organisation (UNESCO) Research Discussion Paper, which found that Black, Indigenous, Jewish, Arab, and lesbian women journalists were affected by online violence most often and with higher severity.[263]

Underreporting: Only one in four women have reported behaviour to the online platforms when online violence had occurred.[264] Therefore, studies recognise that due to underreporting, the existing statistics concerning misogyny and online violence against women are likely a lower representation of the true situation. Further, 78 percent of respondents from this global survey stated that women are often unaware that options exist to report harmful online behaviour.[265]

Prevalence of online violence and hatred on the basis of LGBTQ+ characteristics

Overall prevalence: An online questionnaire completed by 3,731 LGBT young people aged 11–19 in secondary schools and colleagues across the UK uncovered that 40 percent of LGBTQ+ young people have been the target of homophobic, biphobic, or transphobic abuse online. In particular, 58 percent of trans young people have been the target of this form of online abuse.[266] Another study showed that out of 2,538 all-age group participants across various EU member states and the UK, 66 percent of respondents had experienced anti-LGBTQ+ online hate crime and/or hate speech. Further, 20 percent of participants reported that they had experienced online hate speech/crime in over 100 instances in the past five years. This study also confirmed that online abuse is more common for trans victims as opposed to cis-gender LGBTQ+ individuals.[267]

Homophobic, biphobic, and transphobic content is widespread: A significant 97 percent of 3,731 LGBTQ+ young people across secondary schools and colleges in the UK had seen homo-, bi- or transphobic content online in 2017. Furthermore, 43 percent of respondents report seeing this content 'often'.[268]

Prevalence of online hatred and violence on the basis of religion

Antisemitism: The European Commission found alarming trends online from January 2020 to March 2021, including a sevenfold increase in antisemitic postings on French language accounts and a more than thirteenfold increase in antisemitic comments within German channels studied during the pandemic.[269]

The European Union Agency for Fundamental Human Rights reported on the findings of Jewish people across various EU Member States and the UK.193 In relation to online antisemitism it was found that online antisemitism was rated as the most problematic form of antisemitism, and that:[270]

- 89 percent of all respondents considered online antisemitism to be a problem in the country they live in
- 88 percent believed that online antisemitism had increased over the past five years, with most saying that it had increased 'a lot'
- 80 percent identified the internet as the most common forum for negative statements about Jewish people.

A study by the Centre for Countering Digital Hate concluded that the 714 posts containing antisemitic content found on Facebook, Instagram, Twitter, YouTube, and TikTok had generated over 7.3 million impressions.[271] The study also noted 84 percent of posts containing anti-Jewish hate were not acted on by social media companies after our researchers reported them.[272]

> Until recently, the seriousness of online hate speech has not been fully recognised. These statistics prove that activities which unfold in the virtual world should not be ignored.
>
> Dr. Malcom Williams
> Professor, Cardiff University[279]

Islamophobia or anti-Muslim Hatred: Researchers from the Center for Countering Digital Hate (CCDH) conducted a study on social media platforms and identified over 500 posts containing anti-Muslim hatred, which collectively received over 25.5 million views. The posts were found on Facebook, Instagram, TikTok, Twitter, and YouTube during February and March of 2022. Users on Instagram, TikTok, and Twitter were found to be allowed to employ hashtags such as #deathtoislam, #islamiscancer, and #raghead, resulting in at least 1.3 million impressions for the content.

Furthermore, the research found Facebook hosted pages and groups dedicated to promoting anti-Muslim hatred, with a total of 361,922 followers or members across the US, UK, Canada, and Australia. The content of these unaddressed posts included false claims about inherent Muslim violence, dehumanising depictions of Muslims, and racist caricatures. Across Facebook, Instagram, TikTok, Twitter, and YouTube, the researchers discovered 20 posts glorifying the Christchurch massacre terrorist or featuring footage of the attack, with platforms failing to act in 70 percent of the cases. Analysis also found nearly 100 posts referring to conspiracy theories and sites failing to act on 89 percent of such posts. The report noted "the average performance of both platforms owned by Meta shows a failure to act on 89 percent of content… [and] YouTube failed to remove any of the eight videos we reported."[273]

According to a report released by the Institute for Strategic Dialogue (ISD), researchers analysing TikTok have uncovered concerning content promoting extremism. The report highlighted videos on the platform that depicted Muslims as supporters of terrorism, posts that endorse Holocaust denial, and content that supported mass shooters involved in the Christchurch Mosque and Tree of Life Synagogue attacks. These were some among numerous examples of antisemitic content that displayed and incited hatred and violence.[274]

Protestant and Catholic tensions: Global Witness (a human rights group) found as part of research into Facebook's advertising review system that advertisements with hate speech, sectarian slurs, and content inciting violence in Northern Ireland would be approved by Facebook's system.[275] Further, the sheer number of anti-Christian hate crimes reported in Ireland in 2021 (775 incidents and 484 victims) far exceeds the number of reported hate crimes linked to anti-Muslim hate (289 incidents and 453 victims).[276] This statistic relates to hate crime generally. More research would be beneficial to understand the proliferation of online anti-Christian hate crimes.

Relation between online and offline violence: Academics from Cardiff University's HateLab project conducted a study analysing the association between hate speech on Twitter, and racially and religiously aggravated crimes in London. The research found that as the number of 'hate tweets' increased in a specific location, so did the number of related offline crimes. The study suggested that "an algorithm based on their methods could help predict and prevent spikes in crimes against minorities by allocating more resources to specific areas."[277] The research highlighted the connection between online hate victimisation and real-world harm, even in the absence of major events as triggers.[278]

Mahalah's opinion of her Jewish Israeli identity has changed over time in response to anti-Israel stances and antisemitism she perceives in the news and world around her. Mahalah (26–30) was born in Israel and is Jewish.

After recent events in the Israel-Palestine conflict caused new waves of social media engagement with individuals often vocalising anti-Israel or anti-Palestine sentiment, Mahalah felt targeted and troubled by pro-Palestine stances taken by big brands and other social media users. She believed that this presented a simplified, one-sided view of the complicated conflict. Over the course of several months, as the topic circulated on social media, Mahalah began to feel increasingly alienated from wider society. She believes she was made to feel ashamed of her heritage, a process she found very stress-inducing. In the long-term, she is still worried about revealing she is Israeli, especially to her colleagues at her new job.

Revealing Reality, How people are harmed online: Testing a model from a user perspective (2022) Available at: https://www.ofcom.org.uk/__data/assets/pdf_file/0023/244238/How-people-are-harmed-online-testing-a-model-from-auser-perspective.pdf

1 in 2

women had experienced online abuse using digital technology

Women's Aid
Irish national feminist organisation working to prevent and address the impact of domestic violence and abuse[248]

> I'd been bullied in the past so it was just part of my existence. I started getting death threats online after I came out. I told my head of year, but they just told me to come off the Internet. It carried on for years.
>
> Amy, 18
>
> Stonewall, The School Report (2017).
>
> Available at: https://www.stonewall.org.uk/system/files/the_school_report_2017.pdf

5.7.2 Impact of this harm

Promotion of culture of violence: There are a number of studies that have investigated online platforms and the role they can play in increasing technology-facilitated gender-based violence against women (VAW).[280] Incidents have also been investigated by law enforcement agencies to study the incitement of violence using online platforms.[281]

Emotional distress: From 1,662 women who experienced online violence in the past year, 43 percent felt unsafe; 36 percent felt embarrassed; 35 percent experienced emotional harm; 23 percent experienced harm caused to a personal relationship; 17 percent said their families felt unsafe; and 10 percent experienced (offline) physical harm.[282] This is supported by the Plan International (2020) study, which surveyed 14,000 girls and young women from 31 countries. Of the 58 percent that experienced online harassment, 24 percent felt physically unsafe, 42 percent lost self-esteem or self-confidence, 42 percent felt mentally and emotionally stressed, and 18 percent have problems at school.[283] As illustrated in Section 5.5, emotional distress is closely linked to suicidal thoughts, attempts, and self-harm cases.

Changing online behaviour: An Australian e-Safety Commissioner survey of 1,491 women's experiences with online abuse in their working lives found that in response to online abuse, women changed their behaviour from engaging less with platforms to limiting their presence. But women also often described a more subtle self-censorship – pausing before posting a comment, or using a different professional name.[284] This is further supported by Plan International (2020) study, which found that 19 percent of girls harassed frequently said they use social media less – and 12 percent have ceased using it altogether.[285, 286]

Economic and social exclusion: Women and other groups who are disproportionately affected by incitement of hatred and violence online often choose to withdraw or minimise their online presence and activity as a result of being victimised or witnessing others suffer through online hatred and violence (see Section 5.7.4). As a result, the online community can have fewer members of affected demographics as active participants over time, which makes the online world less representative of the population and can also lead to economic exclusion as, for example, women may be discouraged from pursuing careers as journalists or public figures. Feelings of exclusion can also be linked to emotional distress. Research by Economist in 2020 found out of the women surveyed, nearly nine in 10 women restrict their online activity, limiting their access to employment, education, healthcare, and community and seven percent of those surveyed had lost or had to change their jobs due to online violence.[287]

Cultural intolerance and radicalisation: The NSPCC states that experiencing community tension among different groups, having low self-esteem, and a feeling of rejection are among various vulnerability factors that can make a young person more susceptible to radicalisation.[214] These are all emotional states that can be exacerbated by the hatred and violence displayed and incited in online spaces, sometimes resulting in the acceleration of marginalised individuals towards radicalisation. This discussed further in Section 5.9.

> Someone made a Facebook page about me being gay and half of the year liked it.
>
> Edward, 19
>
> Stonewall, The School Report (2017).
>
> Available at: https://www.stonewall.org.uk/system/files/the_school_report_2017.pdf

	Physical	Mental	Moral
Direct impacts	• Self-harm • Physical harm • Destruction of property	• Shame • Self-blame • Depression • Anxiety • Stress • Anger • Sadness • Fear	• Inequality • Avoidance behaviour • Harm caused to personal relationships • Exclusion of people from oppressed demographics
Indirect impacts	• Radicalisation	• Social isolation • Reduced time on social media • Normalisation	• Stifling of human rights • Economic exclusion • Limitation of diversity in online spaces • Cultural intolerance

Table 12: Impact of content promoting hatred and inciting violence on the physical, mental, and moral development of minors

5.7.3 VSP features which can enable harm

User-generated content: VSPs are inherently based on user-generated video content, which often has features for comments to be posted alongside the video content. Online abuse and hate speech that can take place on the grounds of protected characteristics can closely relate to online bullying (see Section 5.3). This can take many forms, such as the experiences of the 3,713 surveyed LGBTQ+ young people residing in the UK:[289]

1. Images, comments, videos, or messages that are private, untrue, mean, or embarrassing (experienced by 30 percent of young people)
2. Threatening messages (experienced by 11 percent of young people)
3. Being filmed or photographed without consent (experienced by six percent of young people)
4. Impersonation (experienced by three percent of young people).[290, 291]

Of 700 reports received by iReport.ie in 2020, racist language was used in 181 instances (54 percent), language about religion in 50 instances (15 percent), and racist symbols or insignia in 36 cases (11 percent).[292]

Recommendation systems and algorithms: Algorithms prioritise engagement, and as such this often leads to the proliferation of hateful content. Engagement intended to criticise the content also makes it more popular and causes it to be more heavily promoted by the algorithm. While TikTok has taken steps to address content that infringes its community guidelines, the Institute for Strategic Dialogue (ISD), emphasised the need for greater transparency and understanding of TikTok's algorithm.218 Facebook whistleblower Frances Haugen stated in 2021 that the platform's systems incentivise users to publish divisive and polarising content.[294] Furthermore, Mark Zuckerberg illustrated in a public note that the closer content gets to being prohibited (if it is in violation of the platform's community guidelines), the more engagement it likely receives.[295]

Targeted advertisements: In June 2021, Global Witness Organisation conducted investigative research to test Facebook's review system for advertisements. As part of the research, political advertisements targeted at individuals in Northern Ireland that contained sectarian language and images were approved by Facebook's systems.[296] Further research conducted by Global Witness Organisation in 2023 also found that advertisements submitted to YouTube, TikTok, and Facebook containing violent, hateful language against LGBTQ+ individuals in Ireland were nearly always approved – 10 such advisements were submitted to each VSP, with Facebook rejecting only two, and TikTok and YouTube approving all of the advertisements for publication.[297]

5.7.4 Specific response measures

Community reporting or flagging

General victim and witness responses:

1. *Reporting:* A survey with 700 respondents across the UK revealed that more than one in four victims did not report their online hate crime abuse to anybody; 44 percent reported it to social media companies; seven percent reported their abuse to the police. In addition, 73 percent of people were unhappy after reporting an incident as no action was taken; 53 percent felt that the incident was not treated seriously; and 20 percent received no response at all. Of the 11 percent who were satisfied with the response they received from the social media platform after reporting and incident, only 6% felt satisfied that the respondent faced consequences.[298]
2. *Not reporting:* Of respondents, 56 percent claimed that the incidents happen too often to tell anyone about it; 55 percent gave the reason that they 'just wanted to forget about it and move on'; while 36 percent were 'afraid that responding in any way would make it worse.' It is also notable that 11 percent did not want judgement for the incident context.[299]
3. *Support:* 42 percent of respondents mentioned the incident to a friend; 20 percent to a partner; and 13 percent mentioned it to other members of their family. Only three percent brought up the incident with a victim support organisation.[300]

Responses among young people:

1. *Reporting:* 30 percent of people in the study encountering online hate speech reported it to the social media platform, website, or app.[301]
2. *Not reporting:* Ofcom's 2017 survey of users found 59 percent had done nothing about or ignored hateful content they have seen online.[302]
3. *Support:* 24 percent of young people encountering online hate speech told a friend, while 22 percent supported the victim by saying something positive.[303]
4. *Educators as source of support:* The majority of teachers took some form of action in relation to their students being involved in hate speech incidents – mostly providing support to the victim through positive words, but also discussing with a colleague, their school principal or someone whose job it was to help young people.[304]

Responses among women:

1. *Blocking contacts:* In one survey, 55 percent of women took this action as a response measure, while a further 24 percent changed their mobile number.[305]
2. *Reporting:* 37 percent of surveyed women reported the grievance with the platform, while only 14 percent reported the behaviour to an offline protective agency. The survey noted, however, that "owing to underreporting, our results may underestimate actual prevalence rates of online violence against women."[306]

Examples of preventative and reactionary practices and strategies beyond reporting

Included below are examples of the following forms of response measures:

1. NGO advice and support for users
2. Promotion of positive content and communities on the VSPs
3. Community reporting or flagging
4. Digital citizenship.

Social, cultural, and educational approaches are needed to tackle the underlying causes of hate speech. Digital citizenship education, including human rights education and media literacy, is crucial in preventing hate speech online. Awareness campaigns, information dissemination, analysis of hate speech, and counter-speech initiatives are common strategies. However, a comprehensive assessment of the impact of educational and cultural responses to hate speech is lacking, emphasising the need for alternatives to legal measures.[307]

Lydia sees a lot of content which she believes has an ageist slant which has changed the way she sees herself in society and impacted her mental health.

Lydia (66–70) lives alone. She has multiple sclerosis and struggles to leave the house. Being unable to socialise in person, Lydia spends a lot of time online using social media and reading news websites. Lydia believes that old people are subtly ostracised in society, in the news and on social media. She used to have a preferred social media platform, but she withdrew from it during the Brexit vote when she felt it became overwhelmed with ageist content.

During Covid she frequently saw old people referred to as 'bed-blockers', 'bed-wetters', and 'senile'. Now, Lydia uses other platforms and websites, but she still sees ageist remarks in the comments on a daily basis. For Lydia, comments like this are indicative of a society which does not have time for older people. This has lowered her self-worth and she feels that people's negative perception of her as an older person in society has added to her depression."

Revealing Reality, How people are harmed online: Testing a model from a user perspective (2022) Available at: https://www.ofcom.org.uk/__data/assets/pdf_file/0023/244238/How-people-are-harmed-online-testing-a-model-from-auser-perspective.pdf

Adil (26–30) received homophobic messages, threats, and abuse after coming out as bisexual on a social media platform. A lot of the comments were from people he knew through the South Asian society at university.

He felt ashamed and worried for his daughter, who was often featured in the abusive messages. He suffered from depression and anxiety following the incident and sought counselling.

Revealing Reality, How people are harmed online: Testing a model from a user perspective (2022) Available at: https://www.ofcom.org.uk/__data/assets/pdf_file/0023/244238/How-people-are-harmed-online-testing-a-model-from-auser-perspective.pdf

Incitement to violence or hatred

Services	Access Now Digital Security Helpline supports women at risk to improve their digital security practices and provides rapid-response emergency assistance for women already under attack. The service, which is available 24 hours a day, seven days a week in eight languages, is monitoring and drawing attention to digital rights during the humanitarian crisis.
	Net Tech Project at the National Network to End Domestic Violence (NNEDV), discusses technology, privacy, and safety in the context of intimate partner violence, sexual assault, and violence against women during COVID-19.
	El Alto, Bolivia under UN Women's Safe Cities and Safe Public Spaces global initiative is developing communications material in a simple and user-friendly format to demonstrate the harmful impact of online violence on women and girls and how to report it during COVID-19.
	The Action Coalition focused on Technology and Innovation under Generation Equality is addressing data collection, prevention, and response of online/ICT (Information and Communication Technology) facilitated violence against women as a priority.
Awareness raising and training	Take Back the Tech! is a campaign around online gender-based violence (GBV) awareness-raising, documentation, and digital safety that has accompanied women, queer, and gender diverse people who have experienced online GBV since 2006.
	#SheTransformsTech is a crowdsourced campaign and global poll that will synthesize stories and input from grassroots women from 100+ countries into a recommendations report for global policymakers and technology companies.
	Tactical Tech Training Curriculum on gender and technology brings a feminist and holistic point of view to privacy and digital security trainings.
	Onlineharassmentfieldmanual.pen.org/, is a digital manual including effective strategies and resources that journalists and advocates can use to address online abuse.
	Internetlab.org.br/en/ is an independent research centre that aims to foster academic debate around issues involving law and technology, especially internet policy.
	GenderIT.org provides gender and ICT analysis informed by African feminists from 18 countries.
	EQUALS Digital Skills Fund is a grassroots fund aimed at supporting digital skills of women and girls across Africa, Asia, and Latin America.
Privacy and the safety of users	Feminist Safety Reboot creates safe online spaces and promotes understanding online GBV.
	Heartmob provides an online support group for affected individuals.
	Jigsaw is undertaking research and technology developments to address violence and harassment online against women in public life.

Preventative practices and strategies aimed at improving women's online safety[308, 309]

5.8 Offences relating to the online identification of victims, suspects, or vulnerable people

This section will focus on the prevalence and impact of content related to the identification of victims, suspects, or vulnerable people (including witnesses) on VSPs. Additionally, it will cover the VSP features that can enable this harm, as well as any specific response measures that can mitigate it.

This will cover multiple offences under Schedule 3 of the Online Safety and Media Regulation Act 2022.

5.8.1 Prevalence and risk of harm

Very few academic studies exist which examine the prevalence of victim, suspect, and vulnerable people identification in online content. As such, this section focuses on notable international cases and examples. We have not identified any Irish examples to include.

Identification of victims

Former Idaho state representative Aaron von Ehlinger and Priscilla Giddings are being sued by Ehlinger's rape victim for disclosing details about her to a media outlet, resulting in her identity being published. Giddings also reshared the article and image identifying the victim on Facebook, further distributing the victim's identity on the VSP.[310] Although this case started with victim identification in the mainstream media, it spilled over into content circulated on VSPs and is therefore covered in this section.

Identification of suspects

Following the murder of four University of Idaho students, users on social media platforms and VSPs started making and spreading content relating to possible suspects of this case. It is reported that the TikTok hashtag 'Idaho murders' and its associated different hashtag versions had thousands of posts and more than one billion views.[311] The online traffic on VSPs relating to this case led many users to wrongfully accuse several individuals through online sleuthing, such as a university professor, the victims' neighbour, and one of the victim's ex-boyfriends. Similar identification and outing cases have occurred for many celebrities both as victims and suspects.[312]

5.8.2 Impact of this harm

Suspects identified before conviction can suffer from serious mental and physical harm, as users on the VSPs jump to conclusions about their guilt. Safety fears, emotional distress, and trauma are just a few of the impacts on suspects outed online.[314] Similarly, victims outed and identified on or through VSP content can suffer from emotional distress, invasion of privacy, unwanted media attention, harassment, and abuse.[315, 316] Additionally, online content relating to the identification of victims, suspects, and vulnerable people undermine law enforcement efforts to find or convict responsible parties, and often taint jury pools.[317, 318]

	Physical	Mental	Moral
Direct impacts	• Potential harm caused to named suspects • Harassment (physical)	• Unwanted media attention • Harassment (mental) and abuse • Invasion of privacy • Emotional distress	• Public perception of guilt of suspects before proven guilty by the courts • Wrongful accusations of individuals • Defamation
Indirect impacts			• Hindering of police investigations and tainting of jury pools

Table 13: Summarising the impact of identification of victims, suspects, or vulnerable people on VSPs, on individuals identified[313]

5.8.3 VSP features which can enable harm

Seamless sharing and minimal content moderation at point of upload enables online victim and suspect identification to take place and be shared widely.

Recommendation systems and algorithms often enhance this through recommending popular or viral content to more people, thereby creating a larger audience for these posts.

5.8.4 Specific response measures

Employing specific response measures to combat this harm is difficult, as this is a niche issue affecting an unknown number of individuals. Generally, VSPs should partner with law enforcement organisations and refine their Terms and Conditions to prevent the identification of victims, suspects, and vulnerable people from happening on these platforms. This should be coupled with quick responses from VSPs when content that contravene these Terms and Conditions is flagged.

Digital citizenship can also play a significant part in educating people around the sharing of such posts, as well as the impacts and consequences of identifying victims and suspects online.

Some of the response measures that can mitigate online content leading to the identification of vulnerable people (including children) are similar to the response measures that can be used against content containing CSAM.

5.9 Online content associated with terrorism

This section will focus on the prevalence and impact of content related to terrorism on VSPs. Additionally, it will cover the VSP features that can enable this harm, as well as any specific response measures that can mitigate it.

Terrorism is the use of intentional violence, or the fear of violence, against civilian populations to achieve political or ideological aims.

Terrorism takes many forms and is used by groups codified into the following categories: jihadist terrorism; right wing terrorism; left wing and anarchist terrorism; and ethno-nationalist and separatist terrorism.[319] Increasingly, online content and social media is allowing international financing and recruitment far beyond geographical boundaries, with video production becoming more professional.

> “
>
> She started speaking online to an American Islamic convert and ISIS recruiter named John Georgelas; he may have pulled on her heartstrings a little bit and she went with it. She was vulnerable, her heart was broken, and she was very naive.
>
> Friend of Lisa Smith, speaking after Lisa travelled from Ireland to Syria to join ISIS[320]

5.9.1 Prevalence and risk of harm

Overall prevalence: The terrorist threat within Ireland is not just from Dissident Republicans (DR), as Europol arrest figures (2019–2021) demonstrate:[325]

- Jihadist terrorism (27)
- Ethno-nationalist and separatist terrorism (13)
- Right wing terrorism (1)
- Left wing and anarchist terrorism (0).

Based on a study of 267 reports used to assess individuals convicted of extremist offences in England and Wales, in the period of 2005 to 2017, radicalisation of extreme offenders involving online aspects has consistently increased (83 percent in 2015–17, 64 percent in 2010–14, 35 percent in 2005–09), during the same time period instances of primarily offline radicalisation has decreased.[326]

Terrorism prevalence on VSPs: According to the European Commission, there are up to 400 platforms hosting terrorist content online, and whereas terrorist content used to be disseminated across larger platforms, it is now being spread more widely over lesser-known sites.[327]

- Europol made over 50,000 decisions on referrals to service providers about terrorist content on their platforms between 2015 and 2018; and the UK's Internet Referral Unit alone identified 300,000 pieces of terrorist content between 2010 and 2018.[328]
- Twitter suspended over 1.2 million accounts for violations of its terms of service in relation to the promotion of terrorism between 2015 and December 2017; and Facebook acted on 1.9 million pieces of ISIS and Al-Qaeda content in the first quarter of 2018.[329, 330]

Livestreams of Terrorist Attacks such as the Buffalo, New York attack237 or the Christchurch, New Zealand attack	Sharing of previous terrorist attacks to promote 'success' stories	Recruitment videos promoting an idyllic lifestyle should people
Radicalisation videos aimed at showing would-be sympathisers that their shortcomings in life are caused by the actions of others	Videos aimed at securing funding for terrorist causes and financing of acts of terrorism from supporters	

Table 14: Examples of content associated with terrorism on VSPs[323,324]

- The UK Home Office stated that **almost all UK terrorist attack planners from 2012 to 2017 downloaded, shared, or consumed content and activity associated with terrorism and extremism online**, showing the risk of harm from this online material – not just in recruitment and funding, but in providing enablers for attacks.[331]

Independent websites: Tech Against Terrorism found that global terrorist and violent extremist actors are running at least 198 websites on the surface web (accessible to the general public) and that these websites attract 1.54 million views per month, with the majority of visits coming from Algeria, Pakistan, the US and the UK.[332] From December 2020 to November 2021, Tech Against Terrorism alerted companies of approximately 11,000 pieces of terrorist content.[333]

Digital propaganda capability: In terms of which groups use an online strategy, jihadist and right wing groups are known to be more sophisticated in their use of online content and campaigns, while DR groups (ethno-nationalist and separatist terrorism) are 'leagues behind' in their use of digital methods.[334] This is mainly because DR members are older and less familiar with online techniques, which has the effect of reducing their recruitment of new members.

5.9.2 Impact of this harm

Existing research on the role played by the internet in radicalisation is scarce, often descriptive, and replete with research gaps. There is an abundance of research on the supply side of online extremist content, rather than how interaction with this content affects radicalisation. Meanwhile, the 'demand'-side of online radicalisation (i.e., how individuals engage with the internet) remains understudied.[336]

Most researchers set out to understand whether terrorists would still have been radicalised had they not assessed material online, and several papers argue that terrorist activities existed before the popularity of the internet and that 'guns don't kill people, people do'. So, while it is not possible to state that online activities guarantee radicalisation, it is largely accepted as truth by policymakers, researchers, and the media that they support it. Given that radicalisation can and does result in the perpetration of terrorist acts, there is potential for exposure to online terrorist content to cause the most severe harm to individuals and society.

Mølmen & Ravndal (2021) argue that to understand the way in which online harms support radicalisation we must first understand the process of: Compensation, Isolation, Facilitation, Echoing, Acceleration, and Action Triggering:[337]

- **Compensation** – Would-be terrorist actors start to compensate for failings in their offline world (societal exclusion, differing world views, grievances) by seeking an alternative online presence where they feel affiliation. This connects offline vulnerabilities to online radicalisation processes in an alternative social environment where the threshold for inclusion is low.

- **Isolation** – Individuals experiencing social alienation often seek alternative social belonging online. Once found, these supporting radicalisation groups and forums often leads to increased withdrawal from offline relationships and alienations from the norms and beliefs in society.
- **Facilitation** – It is widely accepted among researchers that online tools facilitate access to ideological development, as well as operational planning. Interviews with right-wing and jihadi extremists stated that online forms present “key source(s) of information, communication and of propaganda for their extremist beliefs” which allows ‘greater opportunity than offline interactions to confirm existing beliefs. The internet in this case doesn’t drive these beliefs, but it reinforces and justifies existing ones.
- **Echoing** – The most dangerous impact of the internet is that it provides access to a community of like-minded individuals where extreme thoughts and ideas can be exchanged and validated. Further, the sense of anonymity allows individuals to be more extreme and may embolden individuals to express behaviours and attitudes that are otherwise unacceptable, creating a virtual echo chamber that dehumanises the ‘out group’, who become perceived as the enemy.
- **Acceleration** – Access to online material typically accelerates the time taken for someone to become radicalised, as their interactions are more frequent and almost always available. Studies have suggested that the time taken from the first moment of exploration to the final terrorist act could become considerably shorter than average timeframes for offline radicalisation, which takes on average five years.
- **Action Triggering** – The moment when an individual switches from ideology to creating impetus to commit acts of political violence. Scholars agree that action triggering may take place solely online, for example through video or written messages from a terrorist organisation’s leadership.

Two short case studies of radicalisation that occurred within Ireland, one example from jihadi terrorism and one from right wing terrorism, show the direct human impact from this online harm:

Lisa Smith’s story

Jihadi terrorism experienced significant benefits from the use of online platforms in recruitment, financing, and publication of its activities. This was most prevalent in the online activities of ISIS, who used sophisticated video production techniques, incorporating storytelling, music, and lifestyle narrative to recruit over 20,000 fighters to its cause from over 100 different countries. One of these recruits was Lisa Smith, former Irish soldier, who travelled from Ireland to Syria after becoming radicalised via Facebook.

ISIS videos are sickening. They’re also really effective (Vox, 2015). Available at: https://www.vox.com/videos/2015/7/6/8886461/isisvideos-burning
Profile Lisa Smith (Counter Extremist Project, 2019). Available at: https://www.counterextremism.com/extremists/lisa-smith

Mark Wolf’s story

Mark Wolf, a far-right supporter, and paedophile living in Dublin, watched videos of the live stream mass shooting attack on two mosques in Christchurch, New Zealand whilst planning to purchase explosives to conduct his own attack in either Ireland or the UK. He was also caught in possession of firearms parts and guides for 3D printing weapons obtained online. He has purchased weapons parts from overseas. Not only does online radicalisation allow for recruitment, but it places the tools to conduct attacks into the would-be terrorist’s hands.

Far-right sympathiser was buying explosives for terror attack in Ireland (The Irish Times, 2023). Available at: https://www.irishtimes.com/crime-law/2023/04/17/far-rightterror-suspect-was-planning-to-purchase-explosives-in-Ireland-gardai-believe/

In addition to radicalisation and everything this can engender by way of harm to individuals and society, exposure to violent online terrorist content can have the same impact as exposure to other gratuitously violent content. This is illustrated by Damien's story (below) and explored in more detail in Impact of exposure to gratuitous violence (section 5.6.2).

> Damien (51–55) was sexually and physically abused when he was a child. Instances of abuse bring up thoughts and emotions from his past. He has been exposed to ISIS beheadings and other gory human harm videos online from a link that a friend sent him.
>
> A particular ISIS beheading has left a mark on him. He has had flashbacks of it and has instances when he has seen people who look similar in the street and had panic attacks.
>
> Revealing Reality, How people are harmed online: Testing a model from a user perspective (2022) Available at: https://www.ofcom.org.uk/_data/assets/pdf_file/0023/244238/How-people-are-harmed-online-testing-a-model-from-auser-perspective-pdf

5.9.3 VSP features which can enable harm

Several features of VSPs support terrorist use of online mediums, which allow them to communicate their message and enables those searching for it to see it. Note: in some instances, third parties who weren't actively searching for this content inadvertently see it, for example, by the sharing of livestreams.

Anonymity – Certain platforms allow for anonymity, or have a lack of identity verification. This allows multiple accounts to be registered and used to upload content with minimal perceived risk to the offender.

Seamless sharing – Online content can be shared quickly to huge audiences with features such as 'liking' and 'sharing', or other platform-specific terminology. This promotes content far beyond the original poster, exponentially increasing the reach of the content before it is identified. On the other hand, further spread of such materials increase the chances of it being reported.

Transient (disappearing) content – Within closed groups, this type of content makes it harder for moderators to identify the material and deactivate accounts, which remain active for much longer periods of time, sharing material of concern with followers.

Live-streaming – Live-streaming allows the would-be attacker to bypass certain upload filters and creates a sense of live action for those watching, gaining more notoriety for the attacker. This was seen in the Christchurch, New Zealand attacks mentioned in the section above. The notoriety points are supported by only 200 people seeing the footage live, but it being subsequently uploaded 1.5 million times.[338]

High accessibility – Low or no barriers to entry for viewing content (for example, age or identity verification) allow people to view content they shouldn't, either by accident or through feeling they are anonymous.

Minimal content moderation at point of upload – Uploading photos and videos with support for large data files allows the use of high definition and long video productions. Depending on the robustness of upload checkers, these can be uploaded, and seen many times before deletion, and content creators often use various techniques such as code words to circumvent automated checks.

Direct messaging – Supporting functions allow for chat functions, forums, and private messaging, linking distribution of material to communication mechanisms.

Communities – Online platforms have much greater reach than local networks, allowing support nets to be cast much wider to attract sympathisers. Furthermore, by expanding this network globally, would-be terrorists have access to a community that is always online, not needing to wait for the next interaction.

5.9.4 Specific response measures

VSPs have taken action to remove terrorist material from their platforms, as well as inhibit upload in the first place. For the most part this has been in response to government policy. Policy examples include:

- The EU requirement for all social media platforms offering services within the EU to take down terrorist content within one hour after it is identified online.[339]

- The UN Regulation Council's Resolution 2617, passed in December 2021, stressed that member states have the primary responsibility in countering terrorist acts and violent extremism conducive to terrorism. It emphasises that member states need to act co-operatively to prevent and counter the use of information and communication technologies, including the internet, for terrorist purposes. This can include recruitment and incitement to commit terrorise acts, as well as the financing, planning, and preparation of their activities. It specifies that member states should take such action in partnership with the private sector, civil society, and other stakeholders.[340]
- The UK Home Office's UK Interim Code of Practice on Terrorist Content and Activity Online set out five principles in response to the Online Harms White Paper's commitment.[341] These place responsibilities on companies to seek to:
    - Identify and prevent terrorist content and activities from being accessible to UK users on their services
    - Minimise the potential for searches to return results linking to terrorist content and activity
    - Facilitate and participate in industry collaboration to promote holistic and effective approaches to rackling terrorist use of the internet
    - Implement effective user reporting, complaints, and timely redress processes to ensure users are empowered and protected
    - Support investigation and prosecution of individuals for terrorist offences un the UK in line with existing legal frameworks and voluntary reporting measures.
- Europol organised a Referral Action Day (RAD) that targeted online jihadist propaganda on 14 July 2021. This brought together law enforcement authorities from Denmark, Germany, Hungary, Portugal, Slovenia, Spain, and the UK. The aim of the day was to identify material to the service providers with requests to review the flagged content against their terms of service.[342]

Governments are also working alongside companies to develop tools and techniques aimed at detecting and blocking terrorist propaganda, for example:

- Remediation: reporting – In response to policy requiring removal of identified content within one hour, tech companies are building this process of reporting and removal of content into their platforms to meet this timeframe. In an ideal scenario, such content would be blocked before upload.
- Remediation: human scanning – Company and Government organisations scour the internet daily to find terrorist content themselves using labour intensive searches. Since 2010, the UK's Counter Terrorism Internet Referral Unit has identified and worked with companies to remove 310,000 pieces of extremist online material. Meanwhile, Facebook removed 9.4 million pieces of jihadist-related content within three months in 2018.[343]
- Prevention: upload scanning – The UK government partnered with tech company ASI Data Science, to develop a new technology that would automatically detect and block terrorist content online. Testing the tool with ISIS propaganda delivered a 94 percent detection rate of ISIS material, with a 99.995 percent accuracy rating. The tool was designed to be integrated with platforms' video upload process, so that the majority of video propaganda is stopped before it ever reaches the internet. The tool is designed to be used on any platform across a range of video-streaming and download sites in real time.[344]
- Prevention: AI – In addition to scanning uploads to their sites, some platforms use AI to scan content already online that may have escaped upload detection. Examples include:
    - Image matching (checking photo and video uploads against previously removed content)
    - Language understanding (analysing text that has been removed for praising or supporting terrorist organisations)
    - Removing terrorist clusters (using algorithms to work outwards from pages, groups, and posts of profiles that have been identified as supporting terrorism)
    - Employing signals including whether an account has been identified as supporting terrorism, or is friends with a high number of accounts that have been disabled for terrorism
    - Flagging recidivism (detecting new accounts created by repeat offenders)
- Prevention: shared industry database – To stop terrorists 'jumping from platform to platform' some of the largest platforms (Facebook, Twitter,

YouTube, and Microsoft) announced the creation of a shared industry database of unique digital fingerprints (or hashes) that allow member companies to identify and remove matching content across their platforms and block content before it is even posted. The membership has subsequently been expanded to other service providers, including Instagram, Ask.fm, LinkedIn, and Snap.[345]

- Prevention: user terms – Companies are spending more time and attention on ensuring that their terms of use accurately describe how their platforms may be used, and the type of content that can uploaded and shared. Correctly documenting their terms of service will help remove accounts and/or content that violates them.

In October 2022, a terrorist attacked and killed two people outside an LGBTQ+ bar in Slovakia. Aside from posting his manifesto before the attack, the shooter also posted on Twitter and 4chan after the attack, which empowered other users to further post hateful comments towards the LGBTQ+ community.

The Council for Media Services (CMS), which is the media regulator in Slovakia, initially collaborated with platforms such as Facebook, YouTube, and Twitter to mitigate further harm and incitement to violence, but found the platforms' responses to be slow and inadequate. As a result, a report was commissioned which found that there were significant failures by the platforms to address online hate, as well as extremist and terrorist content both pre- and post-attack. The report stated that failures were present in platfworms' content moderation systems and resources, responses, penalisation of repeat Terms of Service offenders, and counter-terrorism policies.[335]

5.10 Online content associated with child sexual abuse

This section will focus on the prevalence and impact of content related to child sexual abuse on VSPs. Additionally, it will cover the VSP features that can enable this harm, as well as any specific response measures that can mitigate it.

Child sexual abuse (CSA) is the involvement of a child (anyone under 18) in sexual activity that he or she does not fully comprehend, is unable to give informed consent to, or for which the child is not developmentally prepared and cannot give consent.'[346]

In addition to the above, some organisations use the term child sexual exploitation for abuse that involves any actual or attempted abuse of position of vulnerability, differential power or trust. Others consider 'exploitation' to be a form of abuse, and so use CSA for both. Child sexual exploitation and abuse (CSEA) online is abuse and exploitation which is partly or entirely facilitated by technology.[347]

Online CSA encompasses a range of harms that fall broadly fall into four categories:

- Offences relating to the viewing, production, and distribution of CSAM
- Crimes regarding the incitement of offences against children
- Live-streaming child sexual exploitation and abuse (transmitting CSA and exploitation in realtime over the internet)
- Online grooming (whereby an individual builds a relationship trust, and emotional connection with a child or young person in order to manipulate, exploit, and abuse them).

Note that the above categories are distilled from the harm types covered in the 2021 Global Threat Assessment of child sexual exploitation and abuse online and do not map onto legal definitions.

CSAM is any visual or audio content of a sexual nature involving a person under 18 years old, whether real or not. Although the OSMR substitutes the term 'child pornography', CSAM is the preferred term in the wider child protection ecosystem, as it is felt to more accurately capture the heinous, illegal nature of sexual violence against children, and to avoid confusion with the legal adult pornography industry.[349]

The definitions above (and below) are taken from the WeProtect Global Alliance 2021 Global Threat Assessment (GTA).[350] Published every two years since 2018, and developed in collaboration with key international actors such as the European Commission, UNICEF, the National Centre for Missing and Exploited Children (NCMEC), INTERPOL, and Europol), the GTA is to date the only authoritative global assessment of the threat of child sexual exploitation and abuse online.

There is potential for all these harms to manifest on a VSP, via:

- Users publishing and/or otherwise sharing CSAM on a platform (including through so-called 'non-malicious' sharing, motivated by outrage or humour)
- Live-streaming of CSEA. This can occur as a manifestation of so-called 'self-generated' sexual material, whereby children ostensibly produce sexual imagery of themselves (through coercion). Live-streaming can also include the transmission of abuse for payment, although this is usually done via private (end-to-end encrypted) video calls.

Other CSEA offences may be committed via the use of secondary functionality, for example:

- Posting comments or images under a video which (either implicitly or explicitly) incites others to commit offences against children, which could include sharing tradecraft
- Posting comments or images under a video in order to seek to build a relationship with a child user for the purpose of grooming them.

There is one notable example of offenders making public, predatory comments under videos of children on a popular VSP , with the platform's recommendation algorithm then suggesting further similar videos based on their watch history.[351] Children also report having been approached by groomers on live-streaming platforms and other online environments.

5.10.1 Prevalence and risk of harm

There is evidence to suggest that online CSA is becoming more prevalent. There has been a sustained rise in reports of child sexual exploitation and abuse online in recent years. NCMEC received 21.7 million reports in 2020, 29.3 million in 2021, and 32 million in 2022.[352] The total reports received in 2022 included 88.3 million files, 37 million of which were videos.[353] To support this further, the Internet Watch Foundation 2022 Annual Report found that there was a 137 percent increase in imagery of boys and a 96 percent increase in imagery of girls, showing a clear increasing trend in CSAM.[354]

There is evidence to suggest that this global trend is replicated in both Europe and Ireland:

- Reports of CSA online where either the imagery or victim(s) were linked to the EU rose from 23,000 in 2010 to more than 725,000 in 2019. Reports indicate that the EU has become the largest host of CSAM globally. The IWF traced 41 percent of CSA URLs to the Netherlands in 2021 (down from 77 percent in 2020).[355]
- In December 2022, Ireland's national centre for combatting illegal content online (Hotline.ie) removed 25 percent more CSAM than in the previous 21 years combined.[356]

'Self-generated' sexual material: Recent years have seen a particularly pronounced rise in so-called 'self-generated' sexual material, whereby children are coerced into producing sexual imagery of themselves. In its 2021 Annual Report, the IWF noted a 168 percent increase from 2020 to 2021 in the proportion of 'actioned web pages' displaying 'self-generated' imagery. It noted the issue arising 'via a growing number of platforms' including live-streaming services.[357]

Underreporting: It is worth noting that the number of reports of CSAM is an imperfect indicator of prevalence, largely because CSA is an underreported crime. Combined with the fact that many online service providers do not detect it, there is a high probability that the actual scale of harm is greater than reports suggest. Estimates of the prevalence of CSA ostensibly support this hypothesis:

- The Council of Europe estimates that in Europe, one in five children experiences some form of sexual violence.[358]
- A global survey conducted by the Economist Impact found that 65 percent of respondents in Europe had experienced at least one online sexual harm during childhood. This was higher than the global average of 54 percent.[359]

Coercion of children to 'selfproduce' sexual material	Sharing and further distributing CSAM online	Online grooming
Sharing tradecraft to help other offenders offend and/or evade detection	Live-streaming child sexual exploitation and abuse	Viewing CSAM online

Table 15: Examples of online content and interactions associated with CSA on VSPs

5.10.2 Impact of this harm

CSA has a damaging impact on individuals, their families, their social circle, and wider society. A recent study by the UK Home Office estimated the cost of contact CSA to exceed £10 billion. This estimate included costs: [360, 361]

- in anticipation of child sexual abuse (expenditure on protective and preventative measures such as costs of education and training)
- as a consequence of child sexual abuse (physical and emotional harms to victims and survivors, lost economic output, and costs to health and victim services)
- in response to child sexual abuse (costs incurred by the police and criminal justice system, as well as the cost of safeguarding victims)

Although this figure is not specific to the cost of online abuse, it is worth noting that online forms are increasingly difficult to disentangle from 'contact' offences: most abuse now has an online element. Furthermore, studies have shown that online CSA has as much of an impact on a child or young person as sexual abuse that takes place offline.[362]

The potential impact of CSA is wide-ranging and severe. In 2019, a team of US psychiatrists reviewed the circumstances of four million victims. They found them to be between two and three times more likely to experience borderline personality disorder, depression, anxiety, post-traumatic stress disorder, and eating disorders. Victims are also 2.5 times more likely to make a suicide attempt.

Identified psychosocial impacts include disruption to relationships, and a higher risk of re-victimisation later in life. Notably, the study also found evidence to suggest that the children of women who have been abused are more likely to be abused themselves. Identified socioeconomic consequences ranged from lower earnings to a higher likelihood of being out of work, or generally unemployed.[363] A summary of physical, mental, and moral impacts of abuse is included in Table 16 below.

The impact such abuse has on a child is not uniform, but will vary depending on factors such as: the child's ability to recognise the abuse and seek support, the circumstances of the abuse (abuse committed by family members is often longer lasting with more severe impacts[364]), and the resources available to help them work through and manage their trauma. Factors such as disabilities, sexual orientation, and pre-existing mental health challenges can multiply the impact of abuse as well as increasing vulnerability to abuse in the first place. Some children will suffer impacts throughout their lives while others may face acute suffering in specific moments, for example when 'triggered'.[365]

In addition to leading to many (if not all) of the same consequences as 'contact' CSA, online CSA can have additional detrimental impact through the re-sharing of abuse imagery, which can prolong suffering and re-traumatise children. In a recent survey by the Canadian Centre for Child Protection, 67 percent of CSAM survivors said the distribution of their images impacted them differently than the hands-on abuse they suffered, because the distribution never ends and the images are permanent.[367]

	Physical	Mental	Moral
Direct impacts	• Difficulties sleeping • Extreme tiredness • Poorer overall health • Gastro-intestinal symptoms • Obesity	• Flashbacks/intrusive thoughts • Problems at school • Panic attacks and anxiety • Eating disorders • Self-harm • Depression • Anxiety	• Self-blame • Low self-esteem
Indirect impacts	• Injuries	• Vulnerability to further abuse/ sexualisation • Fear of images being re-shared, discomfort around cameras	

Table 16: Summarising the impact of CSA online[366]

A study conducted in 2015 by the United Nations Office on Drugs and Crime (UNODC) similarly concluded that repeat sharing "serves to re-victimise and thus further exacerbate the psychological damage to the abused."[368] While some online service providers deploy technologies to automatically detect and remove identified CSA imagery, many do not, and others rely on manual processes which incur significant delays. Of the 5.4 million abuse images detected by Project Arachnid in 2021, only 10 percent were removed after more than 41 days.[369]

5.10.3 VSP features which can enable harm

There are various VSP features which act as risk vectors for CSA:

Live-streaming: Live-streaming functionality is a significant risk factor for CSA because detection relies on proactively monitoring live-streams to detect CSA content in real-time. The summary of industry responses to the first mandatory transparency notices issued by the Australian E-Safety Commissioner concluded: "most providers who were asked did not identify specific steps being taken to identify the abuse of children through live video calls, conferences, or streams."[370] To date Yubo is the only platform to have publicly stated that it proactively monitors all live-streams to detect inappropriate behaviour and intervene in real-time.[371] A survey of Tech Coalition members conducted in 2021 also found that only 22 percent of respondents deployed classifiers to detect child abuse in live-streams.[372]

Once a livestream has ended, there may be no trace of it having taken place, so detection at the point of transmission is usually the only route to identifying crimes and perpetrators. Otherwise, the reliance is on the parties involved to report the incident.

The inherent risk associated with live-streaming is exemplified by the case of Omegle, a livestreaming platform which was attracting 65 million visits each month in January 2021. During a two hour period BBC journalists investigating the platform were paired with 'dozens of under 18s', and connected 'with 12 masturbating men, eight naked males, and seven porn adverts'.

https://www.bbc.co.uk/news/technology-56085499?scrlybrkr=90a20a9a

Minimal content moderation at point of upload: Upload filters are mechanisms by which content is scanned either at the point of upload or prior to publication, providing an opportunity to prevent the publication of illegal or harmful content. Some large platforms do have such measures in place, but with limitations:

1. Some upload filtering is purely for the purposes of identifying and intercepting potential copyright infringement, and does not extend to preventing the publication of child sexual abuse.
2. In respect of child sexual abuse, most upload filtering involves checking content against a database of hashes – unique identifiers of known child sexual abuse imagery. This means that if the video contains newly produced child sexual abuse imagery, it won't be detected.[373]

The general failure to prevent the live-streaming or publication of child sexual abuse imagery on VSPs creates a heavy reliance on detection measures post-publication, such as proactive content moderation and/or on users reporting imagery (by which time harm has already been caused – both to the victim and to the individuals exposed). It also increases the risk and potential scale of re-sharing before detection. An assessment of reports made to NCMEC in October and November 2020 by Facebook and Instagram revealed that copies of just six videos were responsible for more than half of the child exploitative content reported in that time period.[374]

Seamless sharing: By enabling 'one-click' distribution of CSAM, seamless sharing is another VSP feature that contributes to enable and amplify child sexual abuse online. Another feature worth noting is the anonymity granted to many VSP users. The failure of platforms to verify users means that the risks associated with publishing or re-sharing any kind of illegal material are low, hence there is very little to disincentivise such behaviour.

Generative AI: Generative AI algorithms can be used to create new content including audio, code, images, texts, simulations, and videos. Although the generation of images and videos is fairly novel technology, there are already known risks of generative AI relating to child sexual abuse online. These include the evasion or subversion of platform safety mechanisms;289 the production of CSAM;290 or the masking of content to avoid detection. The production of content using generative AI is particularly problematic as this may use real images of abuse as input, re-victimising individuals. Further, the production of such content may complicate law enforcement responses and victim safeguarding, as it may become more difficult to determine whether the victims in the content are real or artificially generated. Recent investigations by journalists have found that generative AI is already being used by nefarious actors to create and sell child sexual abuse images online.[377]

> “
>
> I'm paid for every disgusting show that I will do in front of the computer camera with the customer. And while doing every disgusting show, I lost every bit of my self-esteem to the point where I felt disgusted with myself as well. [...] It's like being trapped in a dark room without any rays of light at all. There's no point in living at all.
>
> Ruby, age 16[348]

5.10.4 Specific response measures

There are technical and wider measures that can be deployed to reduce the risk of online child sexual abuse manifesting on VSPs. Technical measures include:

Content classifiers: Content classifiers would ideally be used to scan content before upload or, failing this, as part of a platforms' proactive content moderation to detect videos featuring child sexual abuse. As highlighted in the 2021 Global Threat Assessment of child sexual exploitation and abuse, the use of such classifiers can increase the need for human moderation due to difficulties estimating the ages of children, and the severity of imagery. However, a range of solutions exist and uptake and accuracy rates are increasing. An example is Google's Content Safety API13, or Thorn's classifier, which Thorn claims is 99 percent accurate.[378]

Proactive live-stream monitoring: As noted by Australia's E-Safety Commissioner, detection of child sexual abuse in live-streams is more technically challenging [than detection on platforms] given the volume of content transmitted. However, it points to two examples of relevant countermeasures:

1. Yubo's proactive live-stream monitoring policy[379]
2. 'SafeToWatch', an on-device tool designed to automatically detect and block the filming and viewing of CSAM. According to its developers, it is the only tool in the world that can detect even uncategorised CSAM in real-time.[380]

Hash-based detection and removal: This is an effective way of removing 'known' (previously detected and reported) CSAM, to prevent (re) sharing and further harm. It involves platforms generating unique identifiers for published content (hashes) and cross-checking these against a number of databases containing other unique identifiers of known CSA content. When a match is made, the content is flagged for reporting and removal. The hashing and hash-matching processes are usually automated. Popular hash-matching tools for video content include PhotoDNA for Video, Google's CSAI Match, and Facebook's TMK+PDQF.[381]

Deploying measures such as content classifiers and hash-matching technologies as part of upload filtering would prevent such content from being published on VSPs in the first instance. Platforms deploying such measures would have a duty to outline how they do so while balancing due regard for fundamental rights such as freedom of expression.

Finally, use of age verification and the verification of children online would not only ensure that the harm caused by child sexual abuse is not exacerbated by children being exposed to imagery, but would also enable appropriate action against users who publish or re-share such content.

Toolkits for children, teachers, educators, parents, and caregivers: Leaving aside the broader elements of the response to CSA, non-technical measures specifically aimed at addressing CSA content on VSPs are limited to news, education, and awareness campaigns. These are generally aimed at informing children and adults about the risks of specific platforms, potentially 'risky' behaviours, and offender tactics. One such example is the IWF's self-generated sexual abuse prevention campaign, which aims to empower girls and "warn parents and carers about the risks posed by online predators targeting children."[382] This is specifically relevant to the sustained trend in self-production of sexual imagery (videos and images), including in live-streaming interactions.

5.11 Online content by which a person's behaviour constitutes harassment or harmful communication

This section will focus on the prevalence and impact of content related to harassment or harmful communication on VSPs. Additionally, it will cover the VSP features that can enable this harm, as well as any specific response measures that can mitigate it.

The type of online content discussed includes content in which individuals either distribute or publish threatening or grossly offensive communication about or to another person, or distribute, publish, or threaten to distribute or publish an intimate image without consent.[383]

The content would need to be such that a reasonable person would realise that the acts would seriously interfere with the other's peace and privacy or causes alarm, distress, or harm.[384, 385]

Harassment violates a person's dignity and creates an intimidating, degrading, humiliating, or offensive environment around the victim. Harassment is often closely related to (cyber) bullying and incitement and display of hatred and violence. Offences pertaining to harassment can also be 'unwanted conduct', which could relate to any of the nine grounds of discrimination mentioned in Section 5.7.

According to Cybersmile, online harassment is a broad term, encompassing various negative experiences online, for example: offensive name calling; purposeful embarrassment; physical threats; sustained harassment; stalking; and sexual harassment.[387] In Irish law, harassment and sexual harassment are well protected in the context of work environments under the 2004 Equality Act amendment to the Employment Quality Act of 1998.[388] The Harassment, Harmful Communications and Related Offences Act of 2020 also makes (the threat of) distribution of intimate imagery and recording thereof without consent (also known as revenge porn) a criminal offence. This is also known as Coco's Law.

The OSMR covers the aforementioned offences in an online context, referring to Coco's Law, as well as the Non-Fatal Offences against the Person Act (1997). Non-fatal offences include persistently following, watching, pestering, besetting, or communicating with another person.[389]

Sharing of intimate imagery without consent	Using messaging or posting functions to repeatedly contact an individual against their will	Sexualising user-generated content through comments
Trickery, or soliciting personal	Upskirting, filming, or photographing under a person's clothes without their consent	Pretending to be someone else online (for example masquerading or catfishing)

Table 17: Examples of harassment or harmful communications on VSPs[390]

5.11.1 Prevalence and risk of harm

Overall cyber harassment: An EU survey of 34,948 over-16s found that in Ireland, 13 percent of all respondents had experienced cyber harassment in the past five years. When broken down on the basis of young people between the ages of 16 and 29, the number rose to 25 percent, and 30 percent of young women specifically. The observation that younger age groups were more likely to experience cyber harassment is present across the survey data of all participating EU countries.[391]

In a UK survey with a sample of 534 respondents, the Victims Commissioner found that Online Harassment was the second most prevalent type of online abuse people experienced, as 45 percent of respondents reportedly experienced this. Further, cyber stalking had been experienced by 33 percent, and 31 percent of respondents had their accounts hacked or controlled.[392]

Women's disproportionate risk: In Ireland, 30 percent of young women aged 16–29 have experienced cyber harassment in the past five years versus only 21 percent of men, and 15 percent of women overall compared to 12 percent of men.[393]

In a global study with more than 14,000 respondents between the ages of 15 and 25, of the girls and young women who have been harassed online, 47 percent had been threatened with physical or sexual violence.[394]

In an online survey conducted across Ireland among young people aged 18–25, 55 percent of young women had experienced stalking and/or harassment. The qualitative segment of this study also showed that online abuse was seen as a key platform for abuse among younger cohorts especially.[395] These figures match global estimates from another survey that gathered data from more than 14,000 interviews, and found that 58 percent of girls had personally experienced some form of online harassment on social media platforms.[396] Figure 3 provides some further examples of the reported forms of harassment and/or harmful behaviour experienced by women.

Sexual harassment: Online sexual harassment can be defined as unwanted sexual conduct that occurs on digital platforms. This can take many forms, which can be classed into the following four categories:[398]

1. *Exploitation, coercion, and threats:* In a study with responses from 3,257 young people across Denmark, Hungary, and the UK aged 13–17, nine percent had received sexual threats online from people within their age group, and 29 percent had witnessed this happening. In the same survey, six percent said that someone used sexual images of them to blackmail them in the past year (2017), while 10 percent of respondents said their boyfriend or girlfriend had pressured them to share nude images in the past year (2017).[399] A 2022 report by Dublin City University found that women were disproportionately the targets of image-based sexual abuse.[400]
2. *Sexualised bullying:* In this same study among European young people, it was found that 25 percent of respondents have had rumours about their sexual behaviour shared online in the past year (2017), while 68 percent stated that girls are judged more harshly than boys. In addition, 31 percent of respondents had seen people their age creating fake profiles of someone to share sexual images, comments,

or messages in the past year, while almost half (48 percent) witnessed other young people sharing personal details of someone who is seen as 'easy'.[401]

3. *Unwanted sexualisation:* In the same study, 24 percent of respondents had received unwanted sexual messages and images in the past year, with girls being significantly more likely to experience this (30 percent) compared to boys (13 percent). Almost a quarter of respondents (24 percent) reported that they had received sexual comments on a photo they posted of themselves in the past year, with girls again being significantly more likely to experience this (26 percent) compared to boys (18 percent). Furthermore, 45 percent of respondents aged 13–17 said they had witnessed people their age editing photos of someone to make them sexual, for example putting their face on a pornographic image or placing sexual emojis over them.[402]

4. *Platform prevalence:* A 2017 study with 4,094 US respondents and 4,321 UK respondents found that over 50 percent of both sample groups indicated that Facebook was the platform on which the most bullying, abuse, or harassment took place. This was followed by Twitter at under 20 percent, YouTube at just over 10 percent, and Snapchat and Instagram averaging below 10 percent.[403]

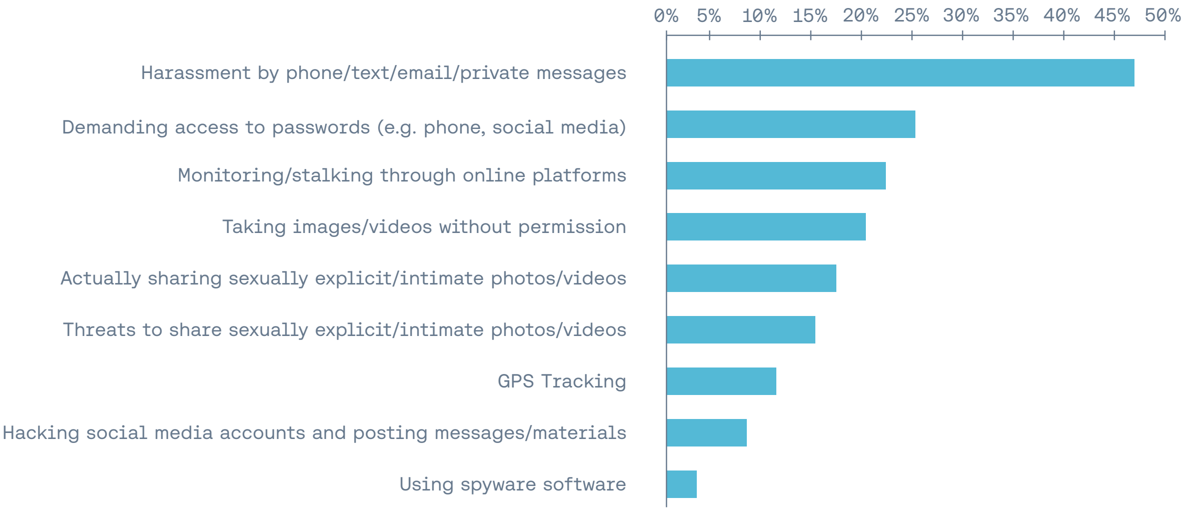


Figure 3: Young women's experiences of online abuse using digital technology[397]

> On Instagram it's easy to step over the line. Over liking photos, sending direct messages, making up multiple accounts. It's hard to block people on Instagram too.
>
> Aoife, associate of online harassment victim[386]

Non-consensual image sharing

Although commonly considered as 'revenge porn', the term inadequately describes it. This is because sharing intimate imagery is not always for 'revenge' purposes, and 'porn' does not grasp the truly damaging nature of such content when it is recorded or shared without consent, making image abuse a more fitting descriptive term.[404] Additionally, the mere threat of sharing intimate imagery can also inflict significant harms on a victim.[405]

Revenge Porn Helpline (RPH) created a report about the individuals that contacted their helpline in 2022 via telephone call or chatbot interaction. Collectively, in 2020 their services accounted for over 9,000 cases.[406]

A Women's Aid survey also showed that 20 percent of young women in Ireland who were subject to online abuse had images or videos taken of them without their consent, 15 percent had been threatened with having intimate photos or videos shared, and 17 percent of young women actually had their intimate images and/or videos shared without their consent.[407]

In a comprehensive study with 3,257 young people across Denmark, Hungary, and the UK aged 13–17, six percent of respondents have had their nude or nearly nude image shared with other people without their permission in the past year, while 41 percent have witnessed this happening to others.[408]

5.11.2 Impact of this harm

A multi-country study carried out in 2020 by Plan International found that one in four girls who have been abused online feel physically unsafe as a result.[409] Young women especially are at a risk of impacts that may be severe, long lasting, and life changing when it comes to intimate relationship abuse, which is often linked to harassment empowered by the use of technology,[410] including VSPs.

From a sample of 534 responses, research from the Victims Commissioner offers the following insights:

- Almost all victims of online abuse reported experiencing some level of harm from the abuse, with only nine percent of people reporting that the abuse did not bother them
- Levels of harm varied depending upon the abuse people experienced
- Victims of intimate image abuse and cyber stalking reported higher levels of harm than victims of other types of abuse.[411]

Norms around gender and masculinity dictate the consequences of image-based sexual abuse, with girls being slut-shamed and victim-blamed, while boys do not tend to face many social repercussions.[412, 413]

Tiziana's story

Tiziana Cantone, aged 31, had taken her own life as a direct result of the distress caused by four individuals who have been under investigation for alleged defamation related to non-consensual image sharing. The distressing video, which Cantone had sent to friends, was further shared without her consent, leading to widespread abuse and mockery directed towards her. Despite winning a 'right to be forgotten' ruling, ordering the removal of the video from websites and search engines, Cantone was still ordered to pay substantial legal costs.

This case has ignited a national conversation in Italy about the need for stronger measures to combat online harassment and protect victims of revenge porn.

CNN, Tiziana Cantone's family calls for justice after suicide over sex tape (2016). Available at: https://edition.cnn.com/2016/09/16/europe/tiziana-cantone-sex-tape-suicide/index.html

Nicole's story

Nicole Fox, aged 21, tragically took her own life due to physical and online bullying, part of which included image sharing. Since then, her mother has campaigned for legislation to better protect people from online abuse which has resulted in Ireland adding to its harassment legislation by introducing the Harassment, Harmful Communications and Related Offences Act 2020 (Coco's Law) in 2021, criminalising the non-consensual distribution of intimate images.

Irish Examiner, Mum fights to have Coco's law against cyberbullying extended across the EU (2023). Available at: https://www.irishexaminer.com/news/arid-41135362.html

Lucy's story

Lucy (16) received unsolicited nude images when she was 13 and only later reflected on the harm this contact had on her. When she was 13, she received a few unsolicited nude pictures from someone she knew from school. In the moment, Lucy admitted that she didn't "pick up on how bad it was", and due to being younger admitted "those kinds of things feel like validation.".

Instead, for Lucy, there was a delayed effect. She said she now realises how "bad" the situation was. It has negatively shaped her perceptions of men and has changed her behaviour on social media by shaping what she posts. She now wants to make sure that things she posts won't attract any unwanted attention from people who might try to do the same thing. She said: "I just don't want that kind of response. I post less now to avoid that stress". She felt these decisions were a small and indirect consequence of the content she saw when she was 13. However, the effect of the unsolicited nudes shared with Lucy shaped her behaviour online over the next few years.

Revealing Reality, How people are harmed online: Testing a model from a user perspective (2022) Available at: https://www.ofcom.org.uk/__data/assets/pdf_file/0023/244238/How-people-are-harmed-online-testing-a-model-from-auser-perspective.pdf

Shruti's story

Shruti (26–30) was a victim of cyber stalking and abuse from her ex-partner. He lived nearby and threatened to visit her house and reveal the details of their romance to her family, a strict conservative family who would not have approved of the relationship.

She felt terrified that he would reveal details of their relationship, and her anxiety affected her performance at work.

Revealing Reality, How people are harmed online: Testing a model from a user perspective (2022) Available at: https://www.ofcom.org.uk/__data/assets/pdf_file/0023/244238/How-people-are-harmed-online-testing-a-model-from-auser-perspective.pdf

	Physical	Mental	Moral
Direct impacts	• Physical harm • Destruction of property • Blackmail (sextortion)	• Shame or self-blame • Loss of self-esteem • Anxiety • Stress • Anger • Sadness • Fear	• Avoidance behaviour • Harm caused to personal relationships
Indirect impacts	• Self-harm	• Social isolation • Reduced time on social media • Depression • Normalisation	• Limitation of diversity in online spaces

Table 18: Impacts of content by which a person's behaviour constitutes harassment or harmful communication[414, 415, 416, 417, 418]

5.11.3 VSP features which can enable harm

Direct messaging: Qualitative research by the charity Refuge in 2022 found that abuse is most likely to take place on the platforms with a direct messaging function. Instagram, Twitter, Youtube, Snapchat, Facebook, and WhatsApp were felt to be most exposed when it comes to intimate relationship abuse.[419] In a 2020 study by Women's Aid in Ireland, Snapchat was noted as particularly problematic, as proof of such abuse can be difficult to obtain as content is only available for a limited amount of time. Screenshotting was also cited as an issue on Snapchat, with perpetrators of online abuse capturing and sharing intimate images without consent.[420]

A study showed found that women perceived all-male WhatsApp groups to be a breeding ground for abusive and 'toxic' behaviour. However, men perceived WhatsApp to be less harmful because a mobile number is required to register and thereby contact the person. Study [X] also found that many social media platforms are being used as tools by partners and ex-partners to abuse and harass young women, leaving them fearful for their safety both on and offline. viewed online abuse as difficult to both address and escape due in part to features including instant messaging and consistent access to content. The effects of this kind of abuse were described as 'draining' and 'exhausting', while at the same time difficult to prove, with no obvious signposts for seeking support or protection.[421]

Anonymity: One of the distinctive features of online harassment is the impunity with which perpetrators feel they can act. The harassment may be visible but the perpetrators themselves can anonymous. There is little redress, as they are unlikely to be caught, let alone penalised.[422] In addition, as creating new and fake accounts is straightforward, there is more opportunity for repeated and longer term abuse.[423]

5.11.4 Specific response measures

NGO advice and support for users: A lack of recognition and labelling of the abuse is seen as a common barrier to victims seeking help and support.[424] Among a study sampling 3,257 young people from the UK, Hungary and Denmark, 28 percent overall did not trust anyone enough to tell about online sexual harassment; while 30 percent also said they 'worried adults would not understand'.[425]

Toolkits for children, teachers, educators, parents, and caregivers: In a study from the UK, Hungary and Denmark, the majority of 3,257 respondents aged 13–17 said they had learned about key topics relating to healthy relationships and online sexual harassment at school. However, many of those did not find this education helpful.[426] Despite this, various studies, including several reports by Dublin City University,[427, 428] have included in their recommendations the need for increased efforts to raise awareness and educate potential victims on consent, digital rights, ethics, and how to report online harassment and other online harms.

With more young men seeking help (84 percent) than women (68 percent), such education could be targeted at young women and girls.[429]

Community reporting or flagging: Most publicly available sources will point victims in the direction of reporting online harassment directly to the VSP.[430] According to Facebook's Community Standards, harassment and other online harms covered in this report are considered unacceptable.[431] However, many surveyed individuals still feel that reporting incidents is ineffective and does not deliver the desired results.[432, 433]

NGO collaboration with VSPs: In the UK, a new platform has been developed to be a central hub for people experiencing online harm to report and remove harmful, abusive, or inappropriate online content. The platform, called Minerva, has been developed by SWGfl, a charity dedicated to empowering the safe and secure use of technology, partnered with the Department for Digital, Culture, Media and Sport (DCMS). The hub will link directly to individual platforms' reporting routes and will maintain a log of reports made, giving those experiencing online harm a simple reporting process and instant access to the reporting avenues they need. The platform will also be developed with law enforcement in mind, as it will create a timeline of events, log reporting activity, and time and date of incidents of abuse, both online and offline.[434]

Projects such as Minerva could help law enforcement effectively manage harmful content online, particularly in countries where questions have been raised about resource constraints inhibiting the police's ability to trace accounts postings at scale.[435] Although an increase in investigative resources may be required regardless, tech-based solutions such as Minerva show promise in being able to help bridge this gap to allow for more effective policing at scale.

Another example is Harassment Manager, an open-source code developed by Jigsaw in partnership with Twitter, allowing users to document and manage abuse targeted at them. Harassment Manager helps users easily identify and document harmful posts, mute or block perpetrators of harassment, and hide harassing replies to their own tweets. Individuals can review tweets based on hashtag, username, keyword, or date, and leverage the Perspective API to detect comments that are most likely to be toxic.[436]

5.12 Online content associated audio-visual commercial communications

This section will focus on the prevalence and impact of content related to audio-visual commercial communications on VSPs. Additionally, it will cover the VSP features that can enable this harm, as well as any specific response measures that can mitigate it.

This includes information conveyed by a media service or relevant online service that is designed to promote, directly or indirectly, the goods, services, or image of the natural or legal entity pursuing an economic activity.[437]

VSPs can be described as online ecosystems that aim to lure in users and keep them engaged. The longer the platforms or content creators can keep users engaged, the longer they are able to sell users' attention to third-party companies advertising on that service.[438]

As per Article 9(1) of the AVMSD, VSPs must comply with obligations about Audio-visual Commercial Communications (ACCs).[440] Generally, these requirements are transposed into national legislation in such a way as to make VSP providers responsible for ensuring their compliance where such commercial communications are marketed, sold or arranged by them. VSPs must also take measures to ensure that when commercial communications are marketed, sold, or arranged by the platform users, such communications meet the same requirements.[441] This is the case for Ireland, as the OSMR 139K (3) sets out.

More detailed measures in national codes or guidance may be required, as set out in the AVMSD and OSMR for the following age-restricted or harmful goods:[442]

- Alcohol[443]
- Tobacco[444]
- Foods and beverages containing nutrients and substances with a nutritional or physiological effect, in particular:
  - Fat
  - Trans-fatty acids
  - Salt or sodium
  - Sugars
  - Infant and follow-on formula.[445]

The UK regulator Ofcom imposes standards on VSPs in respect of two types of advertising on VSPs:[446]

Advertising considered to be marketed, sold, or arranged by a VSP provider: Advertising is considered to be marketed, sold, or arranged by a VSP provider when a VSP provider is involved in making the advertising available on the platform. This may include (but is not limited to) enabling advertisers to buy advertising on their platform either directly or via a third-party, and/or providing tools that enable advertisers to target or optimise the reach of their advert served on the provider's platform.[447]

Advertising not considered to be marketed, sold, or arranged by a VSP provider: Advertising can appear on a platform but not be marketed, sold, or arranged by a VSP provider. For instance, influencer marketing may meet this definition if the VSP has not engaged with the influencer in relation to the advertising. In addition, an advert posted by a brand (in the brand's capacity as a user) on a VSP that appears without any engagement between the brand and the VSP provider would not be considered under the provider's control.[448]

Both adverts that are and are not in control of the VSPs must meet certain requirements, discussed in Section 5.12.4.

> The rapid growth of digital communications and commerce has connected the world in unprecedented ways. Many of these connections come from advertising-supported platforms which provide immense utility to the billions of people who use them. But as the size of the audiences, and the volume of advertising and commerce on these platforms has grown, in turn, bad actors have been attracted to the environments.
>
> These individuals or groups act as advocates for harmful behaviour online, spreading content glorifying harmful behaviours, and at times actively profiting from it. This dynamic is a threat that is too costly for all; people, brands, agencies, and media platforms.
>
> Many industries and organisations have robust responsibility and safety programs around how they source, create, and distribute evaluation of the level of change for the offers and capabilities, in order to make informed decisions and effectively plan for the future of the organisation's products – we must extend this same sensibility into advertising and media given its impact on consumers and society.
>
> World federation of advertisers, Global alliance for responsible media (2020). Available at: https://wfanet.org/leadership/garm/charter

Commercials containing harmful content targeted at users with specific characteristics	Influencers not declaring product placement or sponsorship	Advert campaigns for agerestricted goods reaching or being exposed to underage VSP users
Adverts promoting fraudulent schemes or scams	Adverts promoting unhealthy high fats, sugar, and salt content	

Table 19: Examples of how harmful or misleading audio-visual commercial communications manifest on VSPs

5.12.1 Prevalence and risk of harm

Prevalence online: Over three quarters (78 percent) of consumers in Ireland in 2022 were familiar with the term 'influencer' and almost one third (29 percent) stated that they follow influencers on social media. However, the number following influencers rose to 67 percent for the under-24 age category. Despite this, almost half of influencer advert content is incorrectly not tagged as advertising,[449] and in a survey of 1,000 Irish adults only one in 10 consumers said they trust influencer posts.[450]

VSP reliance of online advertisements: Ofcom's Online Nation Report found that advertising is the main revenue source for most online platforms, with VSPs and other online platforms being primarily reliant on online display advertising.[452]

Prevalence on VSPs: Based on a 2021 survey of 1,958 people in the UK, 27 percent of VSP users declared that they have been exposed to harmful or misleading advertisements while using VSPs in the past three months. Of these people, 16 percent experienced it daily, 35 percent experienced it weekly, and 19 percent experienced it monthly.[452]

Children are at particular risk: Targeted advertising poses serious risk to children through the excessive collection of their data and exposure to age-inappropriate products and services. These risks are exacerbated by children often being unable to recognise commercial content.[453] This is highlighted by a survey of 2,001 children in the UK aged 12–15, which found that 20 percent of respondents strongly disagreed or disagreed with the statement 'I don't mind if websites/social media sites use information about me to decide which adverts they show me'.[454] Research conducted by Ofcom's Children's Media Lives project following a group of 21 children aged 8–18 concluded that most younger children were not aware these online adverts were personalised.[455] A study by the European Commission on the exposure of marketing shows that while only 3.1 percent and 1.7 percent of all advertisements destined for children concerned food and drinks respectively, roughly 90 percent of the 9–17 year-old children had been exposed to advertisements for food and drinks, with parents commenting how these were extremely personalised.[456]

Advertising age-restricted or harmful goods to minors: Alcohol companies have focused considerable investment on content creation within social media, with a significant portion of their advertising budgets and focus now on VSPs. This content often comprises innovative, fun, and creative adverts with video contest, contests, giveaways, and games.[457] A survey of 53 people just above and below the age of 18 found that alcohol brands were more popular on social media with the underage population. In the same study, less than two percent of the posts by brands included messages to 'drink responsibly'.[458] In a study of Instagram posts from 178 popular influencers, researchers found that 19.5 percent of posts containing alcohol showed a clear alcohol brand; however, only a few disclosed this as an advertisement.[459] There is a similar trend among tobacco advertising, with research showing that large tobacco firms are recruiting young influencers to share photos of themselves smoking their brands, frequently at glamorous (tobacco-sponsored) events.

In addition to tobacco and alcohol, in 2019, there were around 15.1 billion impressions on child accounts showing advertisements for products high in fat, sugar, and salt, up from an original estimate of 0.7 billion in 2017.[462]

Lack of available research: There is a lack of literature to determine a sense of the prevalence of both VSP controlled and non-VSP controlled harmful or misleading advertising online.

5.12.2 Impact of this harm

There is a lack of available evidence regarding the impact of both VSP controlled and non-VSP controlled harmful or misleading advertising online.

However, there is some reference to the harm of advertising aimed at influencing children at a subconscious level, termed 'neuromarketing'.[463, 464] Studies have shown that advertising promotes materialistic values in children and can put a strain on the parent-child relationship.[465] This is linked to many children's inability to perform critical thought and effectively evaluate advertisements, making them more vulnerable to exploitation through a seductive allure of marketing attempts.[466]

In terms of influencer marketing, almost a quarter of Irish consumers who purchased a product as a result of an influencer promoting it subsequently felt misled.[467]

“

> You can call it stealth, undercover or guerrilla marketing if you wish. Whatever its name, this is 21st century cigarette advertising that reaches millions of young people around the world.
>
> Robert Kozinets
> USC Anneberg School for Communication and Journalism

5.12.3 VSP features which can enable harm

VSP commercial models are often underpinned by advertising, as expressed by the Institute for Connected Communities:

"Advertising is a fundamental factor influencing platform design/policy decisions, as social media/ Internet platforms are economically motivated to increase site activity to increase advertising revenue. This is achieved by collecting data from users, which is either used to better target advertising or sold to customers and data brokers. As digital platforms are powerful persuasive tools, best practice/regulation protocols are of the utmost importance."[468]

Targeted advertisements: in June 2021, Global Witness Organisation conducted investigative research to test Facebook's review system for advertisements. As part of the research, political advertisements targeted at individuals in Northern Ireland that contained sectarian language and images were approved by Facebook's systems.[469] Further research conducted by Global Witness Organisation in 2023 also found that some advisements submitted to YouTube, TikTok, and Facebook containing violent, hateful language against LGBTQ+ individuals in Ireland were approved. Ten such advisements were submitted to each VSP, with Facebook rejecting only two, and TikTok and YouTube approving all of the advertisements for publication.[470]

Josh (26–30) started investing in cryptocurrency after being encouraged to invest by a former colleague, who had become a 'glamorous' lifestyle influencer by gaining their success from crypto currency and 'crypto-gurus' promoting investment. This friend encouraged Josh to invest in a new crypto currency. Josh did not seek or encounter counter-narratives dissuading or disproving the success of the currency. He invested in a new coin and experienced financial loss as a result, leading to depression and weight gain.

Revealing Reality, How people are harmed online: Testing a model from a user perspective (2022) Available at: https://www.ofcom.org.uk/__data/assets/pdf_file/0023/244238/How-people-are-harmed-online-testing-a-model-from-auser-perspective.pdf

The caller got in touch with the helpline because she recently came across a website via advertising on TikTok and ordered products (razors and accessories) online. Now she has received an open invoice from Klarna (a Swedish financial tech company that provides online financial services), but the website she ordered from has been deleted.

Financial fraud can be difficult for children to identify. It is important to encourage children to consult an adult before making purchases from unknown sellers and to think about the long-term consequences which might include the theft of their financial details.

Better Internet for Kids, Classifying and responding to online risk to children (2023) Available at: https://www.betterinternetforkids.eu/documents/167024/200055/Good+practice+guide+-+Classifying+and+responding+to+online+risk+to+children+-+FINAL+-+February+2023.pdf

5.12.4 Specific response measures

In response to their duties in regulating VSPs, Ofcom has defined advertising requirements for VSPs that fall within three categories: general advertising requirements; prohibited restricted products; and transparency of advertising.[471] These are based on the requirements set out in the AVMSD 2018 Article 9.

Ofcom's approach to regulation is to distinguish between VSP-controlled advertising (advertising that is marketed, sold, or arranged by the VSP provider) and non-VSP-controlled advertising. The regulation designates differing responsibilities to the VSP based on who controls the advertising, as shown in Figure 4.

For both VSP-controlled and non-VSP controlled advertising, the AVMSD 2018 aims to secure the same types of consumer protections through two frameworks.

1. VSP providers are responsible for ensuring compliance with the general advertising requirements and provisions on prohibited and restricted products where the advertising is marketed, sold, or arranged by them. Providers may achieve compliance through a range of approaches, including but not limited to:
   a. ensuring that advertisers and other third-party VSPs engage with and are aware of the relevant requirements (for example, when setting out terms of a contract or during the development of advertising)
   b. ensuring that tools they provide to assist advertisers alert those making use of them to the relevant requirements
   c. taking prompt action to remove or edit advertising that may contravene the requirements, and taking steps to prevent the recurrence of any issues identified with advertising on their platform
   d. determine the appropriate steps they should take to ensure that the advertising they control is compliant with the requirements.
2. Where the advertising is not marketed, sold, or arranged by the VSP provider, VSP providers must take measures as are appropriate to ensure that advertising meets the general advertising requirements, provisions on prohibited and restricted products, and requirements relating to the transparency of advertising.
3. In addition, VSP providers must clearly inform users that a video contains advertising (where they have knowledge of this or it has been declared to them by the uploader), regardless of whether they have marketed, sold, or arranged that advertising.

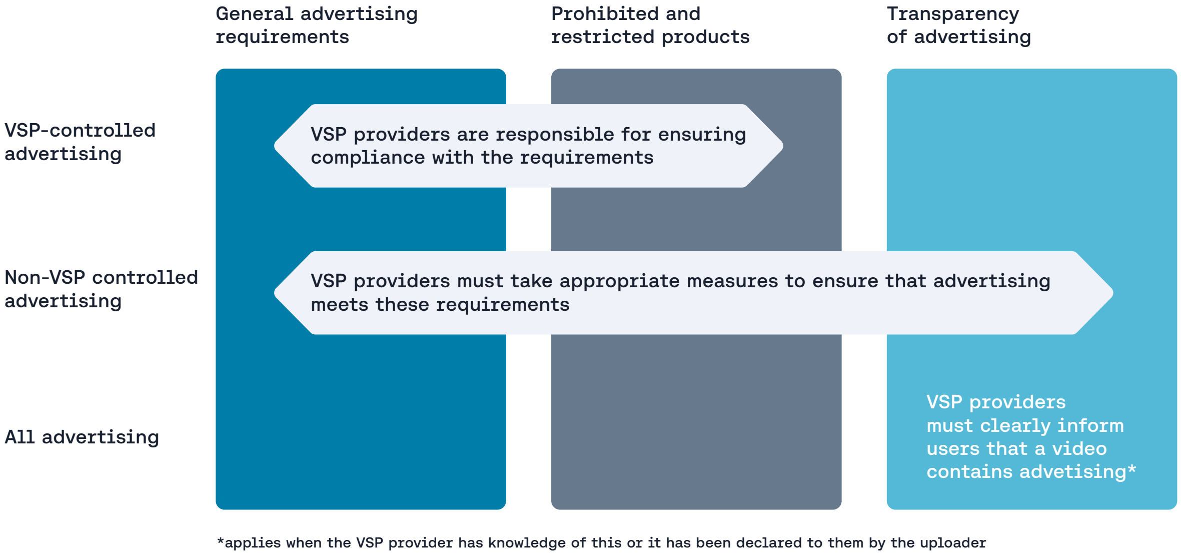

General advertising requirements
Prohibited and restricted products
Transparency of advertising
VSP-controlled advertising
Non-VSP controlled advertising
All advertising
VSP providers are responsible for ensuring compliance with the requirements
VSP providers must take appropriate measures to ensure that advertising meets these requirements
VSP providers must clearly inform users that a video contains advetising*
*applies when the VSP provider has knowledge of this or it has been declared to them by the uploader

Figure 4: Ofcom's overall approach to the regulation of advertising on VSPs[472]

While the requirements given within 'general advertising' and 'prohibited and restricted products' categories are closely aligned to the requirements set out in AVMS Directive 2018, Ofcom has conducted consultation to define the requirements set out in the 'transparency of advertising' category, mandating that VSP providers take the following measures as appropriate to meet requirements relating to the transparency of advertising:

a. make available a functionality for users who upload content to declare the presence of advertising as far as they know or can be reasonably expected to know; and

b. include and apply in the terms and conditions of the service a requirement that users who upload content make use of that functionality as applicable.[473]

Furthermore, Ofcom's 5Rights consultation suggested the inclusion of a requirement that adverts must not use tools that target or extend the reach of the advert to target users under the age of 18.[474]

Due to the nature of user-created content, any implemented measures heavily rely on the collective understanding and awareness of a VSP's user-base. However, currently only 39 percent of users currently feel that VSPs have clear labelling of what is advertising and clear rules for users on how to post advertising content.33 Considering specific areas of confusion from VSP content creators provides context on what type of awareness raising regulation could improve:[476]

1. If the brand being advertised has no/very little control over content (for example, gifted products)
2. The range of permissible declarations within VSP content (for example, specific phrasing/ language, plus at what point within the content the declaration needs to come)
3. If the advertising is only a small fraction of the overall content
4. If rules differ across countries (for example, do UK content creators need to adhere to the same rules as their US peers).

Awareness of rules: For VSP content creators who include advertising within their content, most are broadly aware and supportive of declaring advertising content, and the rules prohibiting advertising for harmful products. However, many lack a clear understanding of the specifics of these rules. In addition, content creators are often unaware of who is responsible for the enforcement of advertising rules, and what the consequences are if these rules are broken.[477, 478] Furthermore, Irish influencers are inconsistent in labelling marketing content, and suggest clearer guidance is needed.[479]

Awareness and engagement with advertising rules can be improved by increasing clarity and understanding about how to declare advertising across the full range of VSP content scenarios. The role of peer-to-peer learning and advice is likely to have an important role to play here, as content creators often gain their understanding of advertising rules from other content creators, rather than through official sources or video sharing platforms.[480, 481]

There are guidelines or best practice documents for brands to follow to protect themselves and their target audience when considering influencer advertising. Some example guidelines are:[482]

1. Find the right creators. Brands should be encouraged to work with creators that do not create inappropriate content and create content that is a 'good fit'.
2. Set clear content guidelines. Create content guidelines to explain the dos and don'ts of creating safe content for children.
3. Disclose paid and gifted relationships. Disclosures must be visible as soon as the user sees the content. Use language clearly understandable by a younger audience.

The UK's Advertising Standards Authority and the Committee of Advertising Practice have produced their own guidelines for influencers, setting out:

1. What the rules are
2. What content should be disclosed
3. Advice around affiliate marketing
4. How to make clear ads are ads
5. Visual examples of best practice
6. What happens if content is not disclosed.[483]

Further guidance documents have been produced in other countries. A full list is contained in Appendix D of the Competition and Consumer Protection Commission (CCPC) report.[484]

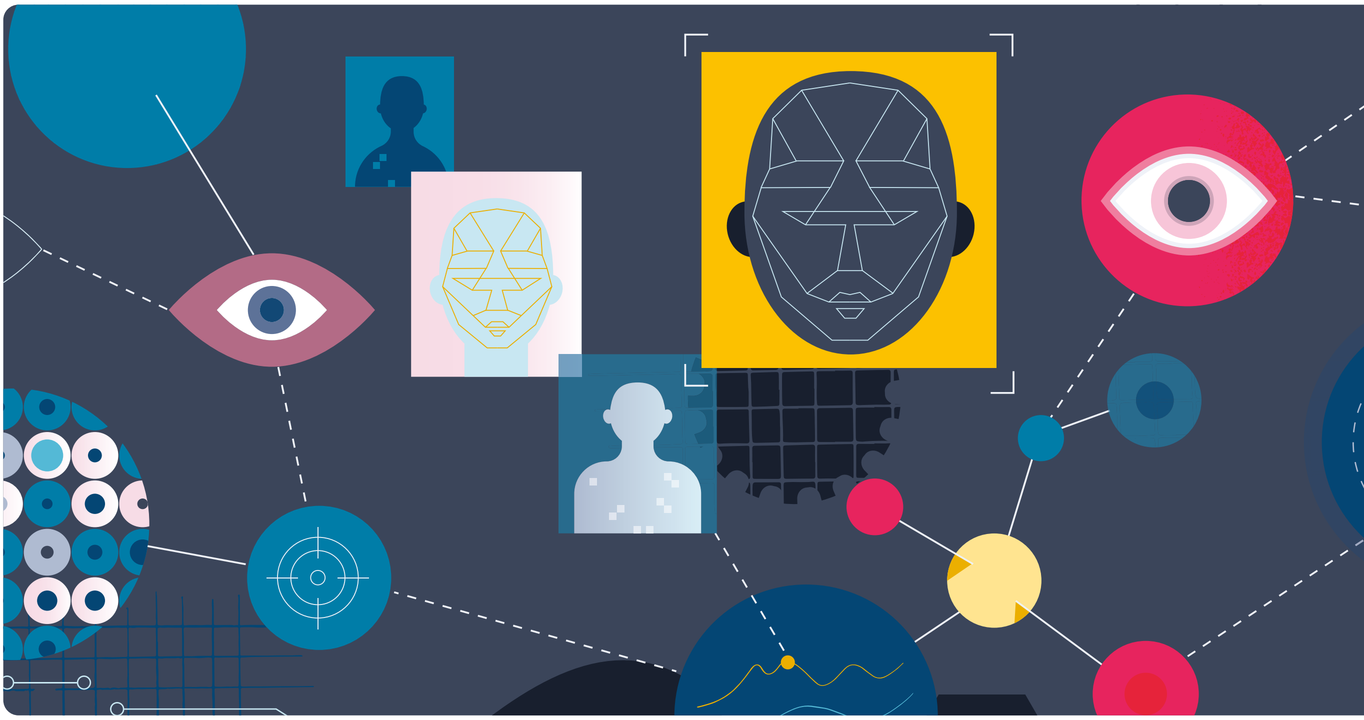
5.13 General features of a VSP that can enable harm

This review has considered how features of VSPs can enable specific harms. Table 20 below provides a summary of features discussed in this report.

General enabling feature	Definition
Anonymity	The absence of personally identifiable information.
Seamless sharing	The ability to easily propagate the distribution of content to a wide audience.
Transient (disappearing) content	Content that expires after a certain amount of time.
Live-streaming	The ability to watch, create, and share videos in real time.
Recommendation systems and algorithms	A feature that filters and recommends items (for example, content, networks, connections, trending topics) based on user preferences and/or past online behaviour. Includes the use of tags to promote content to users with certain interests.
Visual editing	Image or video that is filtered or edited by computer software.
Communities	A collection of users who share similar interests in the content and content creators they engage with.
Influencers	A user high in social standing who has the power to affect their followers' beliefs and purchasing decisions.
Engagement mechanisms	The ability to engage with other users' content to provide feedback (for example, likes, comments, shares).
High accessibility	Low or no barriers to entry for viewing content (for example, age or identity verification).
Targeted advertisements	Advertisements that are specifically aimed at users based on their characteristics, preferences, or past engagements.

Minimal content moderation at point of upload	The frequency or extent that content uploads are checked for harmful material at the point at which they are uploaded to the VSP.
Direct messaging	Enables users to engage in communication privately, including group chats.
Endless scrolling feeds and autoplay	Design that allows for endless scrolling of content that plays automatically, loops upon ending, or automatically moves the user onto the next recommended video.
Lack of identity and age verification	A lack of robust identity and age verification that is hard to circumvent, making accounts traceable to an individual and the platform able to identify the user's age.
Ability to create multiple accounts	The ability for a user to create and use multiple accounts to engage on the VSP.
User-generated content	Any forms of content that have been posted by users on online platforms, tailored, or crafted for a user's feed.
Generative AI	A type of AI system capable of generating text, images, or other media in response to prompts, or acting as a digital virtual 'friend' to the user.

Table 20: Definitions of general features of VSPs that contribute to enabling harm



6

General response measures

Many of the response measures documented within the harm profiles have potentially broad application across a spectrum of online harms.

General response measure	Definition
Toolkits for children, teachers, educators, or parents and caregivers	Resources containing material to educate, upskill, or empower influential people within a minor's life to teach or otherwise educate minors on online harms.
NGO advice and support for users	Resources, campaigns, or helplines for users affected or interested in preventing online harms.
NGO collaboration with VSPs	VSPs consulting or otherwise working with NGOs to develop response measures.
Promotion of positive content and communities on the VSP	Using VSP features to promote content or communities that promotes positive, awareness, or remedial content of online harm.
Terms of service or 'Community Standards'	Terms of service and community standards are determined by individual social media companies, with the intention that users conform to these standards, as well as moderation mechanisms.

Community reporting or flagging	An ability for users to report or flag content as harmful or abusive. This allows VSPs to appropriately respond.
Feedback mechanisms	A process that provides feedback to the user who reported inappropriate or harmful content; particularly on how the report was received, processed, or resolved.
Age verification and the verification of children online	An ability to verify a user's age or identity.
Content rating	Classifying content according to its harm to minors.
Parental controls	Parental intervention includes four overlapping categories: imposing limits or restrictions when online, in terms of time, content, and context; technical (hardware and software) restrictions (for example, filters and parental controls); monitoring online activities either by being a part of the child's online network or tracking/supervising online activities; and actively helping the child to navigate the online environment by having discussion or instruction.
Digital citizenship	Digital citizens can be described as individuals able to use digital tools to create, consume, communicate and engage positively and responsibly with others.
Safety by design	Safety-by-design proposals are a fundamental approach to preventing exposure to harmful content. Two examples of safety-by-design proposals have been identified whereby there are additional measures that online platforms could adopt to ensure a safer online environment for children. There are two main proposed principles: (1) The ICT Coalition for Children Online and (2) Information Commissioners Office (ICO) Age-Appropriate Design Code.
Content classifiers	Technology used to scan content for harmful material before upload or as part of a proactive moderation function.
Proactive livestream monitoring	Human or technology-assisted moderation of live-streams.
Hash-based detection and removal	A method for removing 'known' (previously detected and reported) harmful content.
'Set-to-private' default setting	Ensuring that by default a user's settings are set to pro-privacy.
Anti-recidivism tools	Ensuring users banned for previous reasons cannot re-register with new accounts.

Table 21: Definitions of general response measures that could be leveraged by VSPs[485, 486]

7 Outstanding matters

According to Section 139M of the Online Safety and Media Regulation Act (OSMR), “when preparing an Online Safety Code the Coimisiún must have regard to a number of ‘matters’”. These are detailed in Column 1 of Table 22.

Some of these matters are to some extent covered within Part 1 of this report, in each of the harms profiled. The following sections of this report are intended to complement the Harms Profiles with a more general discussion of these matters, and a focused assessment of items not covered in Part 1.

The discussion of matters (d), (e), and (f) has been combined within Section 7.4 (the rationale for combining the exploration of these matters is provided in the relevant section). The discussion of matter (g) has been separated into a discussion of user rights (Section 7.5) and provider rights (Section 7.6).

The discussion of each matter detailed in Table 22 includes the following elements:

1. Definition (what the ‘matter’ pertains to)
2. Purpose (general relevance to An Coimisiún’s fulfilment of its regulatory duties)
3. Research summary (summary of relevant issues, identified through a targeted literature review)
4. Considerations for regulation (specific complexities/questions relevant to regulatory duties)
5. Potential areas for information-gathering (further opportunities to gather information, not deemed necessary to inform the development of the Online Safety Code, but which may in the future contribute to improve the understanding of pertinent issues).

OSMR S139M matters	Relevant Section
(a): the desirability of services having transparent decision-making processes in relation to content delivery and content moderation	Section 7.1
(b): the impact of automated decision-making on those processes	Section 7.2
(c): the need for any provision to be proportionate having regard to the nature and the scale of the services to which a code applies	Section 7.3
(d): levels of availability of harmful online content on designated online services	Section 7.4
(e): levels of risk of exposure to harmful online content when using designated online services	
(f): levels of risk of harm, and in particular harm to children, from the availability of harmful online content or exposure to it	
(g): the rights of providers of designated online services and of users of those services	Section 7.5 Section 7.6
(f): the e-Commerce compliance strategy prepared under section 139ZF	Not in the scope of this report

Table 22: Details of the matters the Coimisiún shall regard to when preparing its Online Safety Code, and where these matters are discussed within this report

7.1 Transparency

This section provides a summary of available evidence about the desirability of services having transparent decision-making processes in relation to content delivery and content moderation. An interpretation of key terms is provided in Table 23 below.

Key term	Interpretation	Example
• Transparent	• Easy to notice or understand	• N/A
• Content delivery processes	• Processes for the delivery of content (audio, visual, video media) to users of online services	• A stream of content displayed to a user of an online service (also referred to as a 'feed')
• Content moderation processes	• Processes for reviewing online usergenerated content for compliance against policies on what is and what is not permitted to be shared (based on the definition of 'content moderation' adopted by the Trust & Safety Professional Association)	• The detection and removal of CSAM

Table 23: Terminology

Definition

Transparency reporting comprises the regular publishing of information pertaining to an organisation's operations. Reporting can be voluntary, and/or undertaken to comply with regulatory duties. In respect of mandatory reporting, a regulator may have the power to:

- Define the specific information to be included in reports
- Require standardisation to enable comparability over time between different services
- Ensure the information is published in a clear and accessible manner
- Set when or how often a service must publish the information.[352]

Section 139K 4(d) of the OSMR Act stipulates that an Online Safety Code may provide for the making of reports by service providers to the Coimisiún. Such requirements may be used to set a minimum standard that VSPs may then build on by (voluntarily) expanding on the information they provide and/or making reports available to the public (as well as to the regulator). In Australia, for example, in recent years, there has been an increase in voluntary transparency reporting (for example, both Meta[489] and Twitter[490] both have 'Transparency Centres'), and the publication of responses to the first transparency notices issued by Australia's e-Safety Commissioner.[491]

Transparency reports may include information relating to a service's safety practices, content moderation, content recommender algorithms, and action taken against harmful content. Services may further be required to set out transparent complaints processes, publish accessible Terms and Conditions, and submit compliance reports to the regulator outlining their continuous efforts to improve safety.

Requirements to publish reports can sit alongside functions designed to support the broader principle of transparency (for example, facilitating researcher access and requiring or encouraging user empowerment through information, labelling, and other techniques to enable service users to make informed choices).[492]

Purpose

Research for Ofcom, the UK media and telecoms regulator, identified common objectives underpinning the rationale for transparency:[493]

- Exposing malpractice: transparency has been associated with the phrase 'sunlight as disinfectant', and is seen as a valuable principle in exposing malpractice and driving accountability[494]
- Enabling due diligence: transparency has been identified as a crucial element in legislation, setting mandatory due diligence[495]
- Protecting rights: transparency and openness of information is considered fundamental to the realisation and preservation of human rights.[496] The Global Network Initiative's Principles highlight its importance in advancing and protecting the enjoyment of human rights in ICT,[497] and the Children's Research Network believes access to redress and transparency is part of "supporting children to lead positive online lives"[498]
- Promoting consumer agency and competition: transparency can inform and improve consumer choice. This can translate to content services where users may be influenced by perceptions of risk, and the acceptability of actions taken by a particular service. Improving consumer agency can improve the health of markets by driving competition, which in turn can improve price, quality, and innovation for the benefit of users.[499]

7.1.1 Research summary

Transparency requirements are commonly included as core elements of regulatory regimes established to tackle harmful online content. This paper provides high-level findings to enable informed decisions about transparency requirements. By examining existing research and industry practices, this paper has explored the desirability of transparency and its potential impact on fostering accountability, user trust, and regulatory oversight.

The Digital Services Act prescribes new rules that Very Large Online Platforms (VLOPs) will have to follow. Several of these rules pertain to transparency:[500]

- Algorithmic accountability: the European Commission and member states will have access to the algorithms of VLOPs
- Transparency obligations: platforms will allow users to be better informed about how content is recommended to them (recommender systems) and to provide at least one content delivery option not based on profiling
- Independent auditing: VLOPs will have to assess and mitigate systemic risks and be subject to independent audits each year.

Evidencing the impact of transparency

There is a general belief in the benefit of transparency, particularly with respect to enabling better policymaking and facilitating a culture of accountability within digital services. Broadly speaking, transparency within online harms should deliver the following impacts:

1. Help shape An Coimisiún's understanding of what companies are doing to keep users safe
2. Inform the development of An Coimisiún priorities for tackling online harms
3. Provide users with relevant information on the steps providers are taking to address online harms, enabling them to make more informed decisions on which services they use
4. Increase industry accountability and shared understanding between users, civil society, government, and regulators.

Transparency requirements should also incentivise innovation, collaboration, and co-ordination between companies. For example, in relation to the sharing of best practice or specific tools and technologies, in recognition of the fact that the perpetrators of online harms may operate between platforms.[501]

However, current research has limitations. For example:

- Further research would be helpful to confirm that transparency engenders trust in organisations and systems, and that trust differs between how stakeholders perceive a given set of information[502]
- Users' awareness of and engagement with transparency reports is unclear, as is the extent to which people change their behaviour based upon improved disclosure[503]
- There is little comparative analysis about the similarities, differences, and best practices in transparency reporting across different jurisdictions and platforms, to support the development of standardised guidelines and recommendations for effective reporting.

Considerations for regulation

Standardisation: It is currently difficult to enable a standardised format, content, or metrics for transparency reporting,[505] limiting the extent to which reports can be compared effectively across services. Including standardisation within future guidance will increase the ease of report comparison across services. This is important for the understanding of relative performance and progress.

Completeness, privacy, and security: Reports may not be able to provide a comprehensive view of a digital service provider's operations as some information may be subject to legal restrictions, national security concerns, or the need to protect user privacy or commercial interests. Regulators must balance these concerns against the need to enable meaningful transparency and disclosure. User and company rights are discussed in more detail within Section 7.5 and Section 7.6.

Awareness and engagement: The intended audience may be unaware of the existence or significance of transparency reports, or otherwise decline to actively engage with the information. This is a particular challenge when considering young and vulnerable users. Regulators should consider how awareness of transparency reporting can be increased, particularly among young and vulnerable users.

Interpretation: Reports can be 'gamed' (burying critical information in excessive disclosure or presenting data in unhelpful ways), and data can be misinterpreted (an increase in measured illegal content can be an example of improved detection or increasing occurrences of harm). Without contextual information, the meaning of the information can be misconstrued. Standardisation and independent analysis is likely to improve the credibility and accuracy of transparency reporting.[506]

Technical limitations: Transparency, particularly in the context of online safety and algorithmic transparency, can be constrained by a lack of technical understanding. Regulators should look to increase their technical ability to understand and effectively scrutinise transparency reports, as well as having the technical ability to properly inform standardisation guidance.

Holistic view on performance: There is a focus among transparency reports on disclosing aggregated quantitative data on the occurrences of harms and the subsequent moderation interventions. This focus does not provide a holistic view of platform safety, as it fails to improve understanding of the efficacy and impact of content curation, amplification, and moderation.[507] Regulators should consider how to use transparency reports in conjunction with other data sources, such as those shown in Table 30, to obtain a holistic and contextual view on platform performance in reducing online harm.

Proportionality: The requirement and scale of transparency reporting should be considered with regard to Section 7.3, as it may not be proportionate or beneficial to impose the same transparency requirements on every VSP.

A multi-stakeholder Transparency Working Group convened by the UK government in 2020 reached broad consensus on the categories of information that would be included as part of an 'ideal' transparency reporting regime to realise its objectives. Table 24 below provides an overview of these categories, and highlights known examples of transparency reports that include such information.

Network Enforcement Act (NetzDG) of Germany: Transparency requirements for content moderation decisions

In 2017, Germany passed the Network Enforcement Act (Netzwerkdurchsetzungsgesetz, NetzDG) (also called the Facebook Act). The law did not create any new duties for social media platforms but did impose high fines for noncompliance with existing legal obligations. The Act is applicable only to social media networks that have two million or more registered users in Germany. It obligates the covered social media networks to remove content that is 'clearly illegal' within 24 hours after receiving a user complaint.

On June 28 2021, the Act to Amend the Network Enforcement Act entered into force in Germany. The Network Enforcement Act's objective is to combat online hate speech and fake news in social networks. The amendment aims to increase the information content and comparability of social media providers' transparency reports and improve the user-friendliness of the reporting channels for complaints about unlawful content.

Social media networks that receive more than 100 complaints about illegal content in a calendar year are required to publish biannual reports in German on how they deal with these complaints. The amendment requires that more information be included in the transparency reports. Among other things, providers must report if procedures for the automated detection of illegal content are used and, if yes, how they work. In particular, providers must report what training data for the system is used and what procedures for quality assurance or evaluation are in place. In addition, reports must further subdivide the numbers of complaints according to the amount of time it took to remove the flagged content (within 24 hours, within 48 hours, within a week, or at a later date). Information on the new appeals procedure must be added to the transparency report, meaning the number of appeals and the number of cases in which the original decision was revised.

The NetzDG transparency requirements have offered a starting point to enable more robust research regarding how moderation practices affect the manifestation of online harms, and are an aspect of the law that has received universal support.[504]

https://www.loc.gov/item/global-legal-monitor/2021-07-06/germany-network-enforcement-act-amended-to-better-fight-onlinehate-speech.

Information category	Examples of reporting where such information is included (non-exhaustive)
Evidence of effective enforcement of the company's own relevant Terms and Conditions	• Meta's online transparency reporting centre (https://transparency.fb.com/en-gb/) • YouTube's community guidelines enforcement (https://transparencyreport.google.com/youtubepolicy/removals?hl=en_GB) • TikTok Transparency Centre (https://www.tiktok.com/transparency/en-gb/) • LinkedIn Transparency Centre (https://about.linkedin.com/transparency) • Snapchat's Transparency Report 2022 (Snapchat Transparency Report | Snapchat Transparency)
Processes that the company has in place for reporting illegal and harmful content and behaviour, the number of reports received, and how many of those reports led to action	• Australian Government eSafety Commissioner's Basic Online Safety Expectations 2022 (Section 5.7) (https://www.esafety.gov.au/sites/default/files/2022-12/BOSE%20transparency%20report%20Dec%202022.pdf) • YouTube's community guidelines enforcement (https://transparencyreport.google.com/youtubepolicy/removals?hl=en_GB)
Proactive use of technological tools, where appropriate, to identify, flag, block, or remove illegal or harmful content	• Australian Government eSafety Commissioner's Basic Online Safety Expectations 2022 (Section 5.1 to 5.6) (https://www.esafety.gov.au/sites/default/files/2022-12/BOSE%20transparency%20report%20Dec%202022.pdf)
Measures and safeguards in place to uphold and protect fundamental rights; ensuring decisions to remove content, block, and/or delete accounts are well founded (especially when automated tools are used), and that users have an effective route of appeal	• No substantial publicly available example found
Understanding how companies are engaging and co-operating with law enforcement and other relevant government agencies, regulatory bodies, and public agencies	• Meta's information for law enforcement authorities (https://about.meta.com/actions/safety/audiences/law/guidelines) • No substantial publicly available example of measuring cooperation between VSPs and law enforcement found
Details of investment to support user education and awareness of online harms, including through collaboration with civil society, small and medium-sized enterprises, and other companies	• No substantial publicly available example found

Table 24: Information categories that would form part of an 'ideal' transparency regime, and examples of current reporting where such data/insight is included[508]

Summary

Regulators need to consider how to create meaningful transparency. They must avoid creating ‘tick-box’ compliance exercises, and focus on ensuring services provide honest and genuine insights. Requiring information that is designed for specific audiences may also improve engagement and user agency. Rather than treating transparency as an abstract concept, regulators should design their approach to target specific objectives. They must ensure information is accessible, accurate, manageable, and verifiable, to support those objectives.[509]

7.1.2 Potential areas for information gathering

Transparency requirements are included within many regulatory frameworks, such as the OSMR, the Online Safety Bill, and Australia’s Online Safety Act. Given this, further data and evidence about the benefits of transparency could support the proportionality of such requirements, as this emerging area of regulation matures.[510, 511, 512, 513] While this report has presented relevant insights, there is scope for further information-gathering to inform the future work of An Coimisiún. Some of these discretionary areas for information-gathering are detailed in Table 25.

The regulation of online services is an emerging area for regulators throughout the world, with greater transparency a key consideration in public discourse.

A	Information that can be independently ascertained as a user of the service
B	Information that can be ascertained from desk-based research using publicly available, third-party sources
C	Information that would need to be requested from the specific service provider
D	Information that can be obtained through the commissioning of new research

(a): The desirability of services having transparent decisionmaking processes in relation to content delivery and content moderation	Suggested information	A	B	C	D
Further evidence that transparency engenders trust in organisations and systems.	Attitudinal surveys				X
Research about user engagement with, and awareness of, transparency; and its impact on user behaviour/ perceptions	Attitudinal surveys				X
Comparative analysis about differences and best practices in transparency reporting across different jurisdictions and platforms	Research that looks at requirements, implementation, and impact – both on platform and user behaviours		X		

Table 25: Possible areas for information gathering with respect to the desirability of transparent decision making processes

7.2 Automated decision-making

Definition

This section provides a summary of available evidence of the impact of automated decision-making in relation to content delivery and content moderation processes used by online services. Interpretation of key terms is provided in Table 26 below.

Part 1 of this report details evidence about the impact of automated decision-making, with reference to the specific harms profiled:

- Content moderation policies and processes are identified as a VSP feature that can enable harm in respect of six of the 10 harms profiled in this report, by either being minimal, inconsistent, or not proactive enough (not occurring at the point of upload)
- Recommender algorithms, a primary method of content delivery, are identified as a VSP feature that can enable harm in respect of seven of the 10 harms profiled in this report.

These references are not repeated here in full but have informed this summary assessment, and should therefore be factored alongside the evidence below when considering this matter.

Purpose

Automated decision-making for content moderation is used by many online service providers because it is quicker, more scalable, and more consistent than the alternative.[514] In respect of the most egregious online content, it can also reduce the requirement for human moderation, thereby limiting staff exposure to harmful material. Automated content delivery involves presenting automatically curated 'feeds' to users to maximise engagement, in lieu of generic landing pages. Facebook unveiled the first 'feed' in 2006.[515] It has since become the principal mechanism for online content delivery.

7.2.1 Research summary

The impact of automated decision-making on content moderation

Automated decision-making for content moderation is critical for large and medium-sized platforms. Reliance on human moderation would arguably have prevented most from reaching their current size, given the huge volumes of user-generated data they process. As an illustrative example, over 300 million photos are uploaded to Facebook every day.[518] In addition to scalability, automation increases the speed of moderation, and reduces human moderators' exposure to harmful content. It also offers the possibility of making users safer 'in the moment', through real-time automated moderation[519] (for example, to prevent exposure to suspected CSA or terrorist content).

The above benefits notwithstanding, it is important to also understand the limitations of automated content moderation. Most are underpinned by machine learning techniques that are used to detect, flag, and in some instances remove potentially new instances of egregious content. But it is posited that most automated moderation tools are simply undertaking 'a sophisticated version of pattern matching', whereby new content is compared to a database of known examples of previously identified harmful or illegal content. This is limited in the sense that it involves simply replicating past judgements and enactment of platform policies, without factoring that they should actually adapt over time, because societal context and language are in constant flux.[520] An Online Harms Feasibility Study commissioned by the UK government similarly highlighted that AI "is difficult to update over time as expressions of online hate change". The study additionally revealed that AI "performs poorly on video, memes and audio compared to text."[521] The inability of automated content moderation to factor wider context into decision-making reveals why humans will continue to play an important

Key term	Interpretation
Content delivery processes	Processes for the delivery of content (audio, visual, video media) to users of online services, also referred to as 'recommender algorithms'.
Content moderation processes	Processes for reviewing online user-generated content for compliance against policies on what is and what is not permitted to be shared (based on the definition of 'content moderation' adopted by the Trust & Safety Professional Association).[516]
Automated decision-making	The process of making a decision by automated means without any human involvement (based on the definition used by the UK Information Commissioner's Office).[517]

Table 26: Terminology

role in content moderation processes going forward. This was highlighted during the COVID-19 pandemic, when most large platforms saw an uptick in errors when human moderators were sent home.[522]

The potential negative impact of content moderation is tied to the above limitations, specifically the fact that the main inputs to decision-making are a platform's past determinations. Platform policies on content are not objective or based on universally agreed standards. Exploring the example of hate speech, identifying content as hate speech "is a social and performative assertion" rather than an act of classification and one which, undoubtedly, "will be disagreed with".[523] Furthermore, there is evidence to suggest that the rules underlying automated content moderation replicate the inherent and structural biases governing society at large:

- In one study, researchers found that tweets written in African American English commonly spoken by Black Americans were up to twice more likely to be flagged as offensive compared to others. Using a dataset of 155,800 tweets, another study found a similar widespread racial bias against Black speeches.[524]

In a 2020 paper on content moderation, the author Tarleton Gillespie stated: "statistical accuracy often lays the burden of error on underserved, disenfranchised, and minority groups".[525] He went on to cite "the margin of error typically lands on the marginal: who these tools over-identify, or fail to protect, is rarely random."[526] Eugenia Sapera posits that automated content moderation not only replicates but reproduces racism: relevant individuals are made into "passive recipients of AI moderation with no significant input in the decision-making processes", and then further oppressed by being hired into low paid (or unpaid) content moderation roles where their work is used to train the very algorithms that perpetuate their exploitation, additionally by gradually reducing their own roles.[527]

The issue of racism points to a broader negative impact of content moderation – which is that it can impose bias and discrimination on a large scale, while doing nothing to address the social and structural causes of online hate and abuse. As Siapera argues in her 2023 article: "These are social problems, and they have to be addressed holistically. If you try to control this problem at the level of circulation, it's just this never-ending whack-a-mole kind of game."[528] This is illustrated by the problem of online CSAM, the detected volume of which continues to increase year-on-year.[529]

The impact of automated decision-making on content delivery

Automated decision-making in content delivery is experienced by users as the 'feed' they encounter when they log on to a platform. Feeds are either algorithmic ('recommender systems'), or chronological. The latter category is more straightforward as it comprises a reel of content organised by the time-stamp. Algorithmic feeds involve analysis of a potentially broad range of factors to compile a list of the items of content that are judged to have the highest chance of keeping a user engaged. Factors weighed in the (automated) decision-making for such content delivery could include: a user's past preferences, 'trending' topics (content proving to be of interest to a large group of users), and demographic information about the user (for example, their age, location, job). The objective is to maximise engagement, for commercial purposes: the longer a user stays online, the more ads they will view and potentially engage with. Users who visit frequently and for long periods can also be classed as 'active users': this is a key metric used to set advertising rates and measure success.[530]

The success of the strategy is ostensibly supported by statistics on social media usage. A 2019 report by Commission for Communications Regulation (ComReg) found that Irish people spend 4.5 hours on their smartphones daily, including an average of 64 minutes on various platforms.[531] Second to watching television, social media use is the activity for which most people in Ireland connect to the internet.[532] There is mounting evidence of the harmful effects of social media use, which has led to suggestions that excessive use should be considered "a distinct form of behavioural addiction" that can lead to everything from reduced sleep[533] to "higher levels of social anxiety and depressive symptoms."[534] A recent article from the Harvard Business Review articulates why feeds can create such an addictive experience:

"Platforms are designed to trap viewers in a social media rabbit hole: they offer bite-sized content that makes it easy to quickly consume several videos or posts in a row, they often automatically suggest similar content, and many of them even automatically start playing similar videos, reducing the potential for interruptions. While presenting users with engaging content isn't necessarily a bad thing, the accessibility of this media is exactly what makes it so hard for users to break free from the rabbit hole and get back to whatever they were working on."[535]

Particularly in respect of children, excess screen time has been linked to multiple adverse effects, including on overall development.[536] The issue of excess screen time was not caused by automated content moderation: it results in large part from the increasingly digitalised nature of the world we inhabit. But 'feeds' – and the strategy behind them – are undoubtedly a contributing factor.

In addition to these broader effects at the societal level, automated content delivery can have a more nefarious impact. In 2020, YouTube came under criticism when it was discovered that suspected paedophiles were using the service to view, discuss, and share videos of young children. The videos themselves were innocuous, but were attracting inappropriate comments. A clue to the nefarious nature of the viewing audience was the fact that "typically a child's videos might have 60 or 200 views, while those that interest paedophiles suddenly ramp up into the hundreds of thousands." The issue was blamed on YouTube's content recommender system, which was in this instance helping individuals with a suspected sexual interest in children to perpetually discover and view similar content.[537]

In 2022, the Centre for Countering Digital Hate published a report of its research into TikTok's algorithm and the promotion of self-harm and eating disorder content on the platform. As part of the research, the Centre created 'standard' and 'vulnerable' accounts, and the study found that vulnerable accounts "received 12 times more recommendations for self-harm and suicide videos that the standard accounts."[538] The study also noted that "TikTok accounts established with the phrase 'loseweight' in their name received significantly more recommendations for eating disorder and self-harm content."[539]

There are other examples of the impact of recommender systems detailed throughout the harm profiles in Section 5, such as:

- The proliferation of an online video of an unprovoked assault on a teenager (Section 5.3.3)
- The significant role of recommender systems in the increased engagement of vulnerable users with pro-eating disorder content (Section 5.4.3)
- The negative effects of self-harm and depression-related online content recommended to young people and the contribution that content can have to physical self-harm and suicide (Section 5.5.2)
- The responsiveness and rate of harmful content exposure by recommendation algorithms for vulnerable accounts (Section 5.5.3)
- The role of recommendation systems in the consumption of sexually explicit or gratuitous violence content by minors (Section 5.6.3)

- The phenomenon referred to as algorithmic hate, and the examples of critical engagements with hate content adding to the popularity of the content and therefore it's promotion by algorithms (Section 5.7)
- The creation of echo-chambers and resulting acceleration of users to commit terrorist acts (Section 5.9.2).

The problem underlying this and other examples of the online 'rabbit hole'[540] is the way in which content delivery algorithms limit exposure to alternative content and viewpoints. As highlighted in a recent paper published by the UK government on behalf of its four digital regulators, while this will result some users only ever seeing innocuous content, for others recommender systems "can result in people being repeatedly shown content that is misleading or damaging, such as antivaxx conspiracy theories, or even violent content." The paper links the issue to the creation of 'toxic online environments', which potentially lead to real world harm and can affect both individuals and society.[541]

The above alludes to a linked, broader issue with automated content delivery: the suggestion that it lacks sophistication, and therefore integrity. Tech journalist Karen Hao has argued: "The machine-learning models that maximise engagement also favour controversy, misinformation, and extremism: put simply, people just like outrageous stuff." Hao evidences her argument by citing testimony provided to the US Senate on the impact of Facebook by former product manager Frances Haugen. In her opening statement, Haugen asserted that Facebook's products "harm children, stoke division, and weaken our democracy." She went on to state: "Facebook... knows... that engagement-based ranking is dangerous without integrity and security systems... in places like Ethiopia it is literally fanning ethnic violence." Haugen argued that a contributing issue in the situation in Ethiopia is Facebook's uneven coverage of languages, which limits detection of harmful content in regions where English is not spoken, thereby worsening the situation in those places.[542]

Finally, like automated content moderation, there is evidence to suggest that automated content delivery can also be discriminatory. Although difficult to confirm, it is suspected that demographic information about a user is a key 'input' to some content recommender algorithms. In such cases, the content that users are exposed to would to some extent be based on analysis of sensitive or protected characteristics – individual acts of discrimination that are prohibited in other contexts. The previously cited UK government paper voiced concern that "a number of algorithmic systems have been shown to produce biased or discriminatory results" based on inherently flawed and prejudicial 'feedback loops'. The UK government paper cited the example of predictive policing models trained on arrest data provided by police forces where discrimination against Black and ethnic minority individuals is a historical issue. It is arguable that at the societal level, similar feedback loops for content delivery could replicate and reproduce discrimination by systemically limiting the exposure of certain groups to certain types of content and, conversely, by 'flooding' the same groups with certain imagery or information.[543]

Considerations for regulation

Adherence to global frameworks and standards for ethical digital service delivery: Various guidelines exist to promote ethical content moderation, in order to ensure that fundamental human rights (such as the right to not be discriminated against) are not compromised by the delivery of digital services. The Online Safety Code might consider how to ensure alignment with relevant frameworks and, secondarily, how platforms' adherence to and alignment to such guidelines may be monitored under the regime. Relevant standards include:

- The Digital Ireland Framework, which outlines – among other things – Ireland's commitment to an ethical approach to shaping the role of AI in Irish lives[544]
- The European Commission's declaration on digital rights and principles[545]
- United Nations draft recommendation on the ethics of AI online[546]
- The Santa Clara Principles on transparency and accountability in content moderation.[547]

Use of measures to manage other risks associated with automated content delivery: There are many features and mechanisms platforms can implement to reduce and manage the risks associated with the types of content users are exposed to, for example:

- Features to encourage users to monitor their screen time. For example, TikTok has recently limited the default screen time on its app to 60 minutes for those under the age of 19[548]

- Parental controls, which parents and guardians can also use to limit screen time and adjust content recommender preferences
- 'User empowerment' features, which increase a user's control over the content they are recommended and exposed to (a requirement included in the UK's Online Safety Bill).[549]

Ofcom-commissioned evaluation of recommender systems in relation to illegal and harmful content: In July 2023, Ofcom published a report authored by Pattern Analytics & Intelligence (affiliated with Oxford University) that assessed methods used on user-to-user services[550] to recommend content. The report covered four key areas:

1. How recommender systems work and associated design choices: Collaborative filtering and content-based approaches are explained.
2. An inventory of evaluation methods to understand the impact of design choices: looking at questions on baseline decisions, process decisions, and definitional decisions.
3. Assessment of these evaluation methods: including general principles for good evaluation methods and metrics that can be used to compare assessment method elements.
4. Best practice guidance in evaluating recommender systems is detailed for service providers: The guidance suggests combining several different overall evaluation methods and that platforms undertake both formative and summative evaluation. It is noted that consideration should be given to the size, nature, sector and resources of the service provider.[551]

By banning the targeting of advertisements at children, the Digital Services Act will go some way to curbing the commercial incentives to maximise user engagement potentially at the cost of safety, at least in respect of minors. Monitoring platforms' compliance with this rule will be key to curbing the potential negative impact of content moderation and content delivery algorithms on young people.[552]

7.2.2 Potential areas for information gathering

Table 27 below provides a summary of the types of information that might be sought and analysed to further explore the impact of automated content moderation and content delivery. Most of this information is held by the platforms themselves.

Information from platforms is crucial to better understand the impact (both positive and negative) of automated decision-making on content moderation and content delivery processes, but is usually commercially sensitive because it can reveal specific methods used by platforms to maintain and maximise user engagement. Limited insight is provided by some platforms by way of transparency reporting, but in the absence of proactive disclosure the primary route to gathering detailed information would be via issuing information requests. It is worth noting that algorithmic transparency will be a requirement for Very Large Online Platforms (VLOPs) under the Digital Services Act.

Otherwise, publicly available platform documents – such as service terms and conditions, and community guidelines – can be examined to understand the principles behind content moderation and content delivery, even if the detail of processes is not outlined. Third party publications that assess platform approaches to content moderation and content delivery may also include relevant information.

A	Information that can be independently ascertained as a user of the service
B	Information that can be ascertained from desk-based research using publicly available, third-party sources
C	Information that would need to be requested from the specific service provider
D	Information that can be obtained through the commissioning of new research

(b) The impact of automated decision-making on content moderation and content delivery processes	Information to assess	A	B	C	D
Evidence about the measures platforms take to ensure that content moderation and content delivery processes are ethical and do not contribute to replicate or multiply harm, and evidence of their impact.	Adherence to global frameworks and standards for ethical service delivery			X	
	Implementation of measures to manage risks associated with automated content delivery and moderation, such as: • 'User empowerment' features that grant a degree of control over the content a user is exposed to • Parental control mechanisms Screen-time monitoring/ default limits	X			
	Compliance with laws regarding the use of user data, specifically pertaining to protected characteristics	X	X	X	
	Statistics regarding user screen-time (separately for minors and adults, if relevant)			X	
	Information regarding the 'inputs' to recommender algorithms for content delivery			X	
	Information regarding inputs to automated content moderation (for example, details of child abuse hash-lists in use)			X	
	Content moderation standards and guidelines in each of the languages in which the platform has a market presence.			X	

Table 27: Possible areas for information gathering with respect to the impact of automated decision-making on content moderation and content delivery

7.3 Proportionality

This section provides a summary of available research, considering the need for any regulatory provision to be proportionate and have regard to the nature and the scale of the services to which a regulatory code applies.

Definition

For the purpose of this report, 'proportionality' is considered the need to tailor a regulatory approach, or requirements (such as a code), to ensure obligations are commensurate to the nature and scale of a particular service.

The nature and scale of a video-sharing platform service can be defined based on analysis of several factors:

- The nature of a service:
  - The type of users the service is aimed at (business users, professionals, children, adults)
  - The average age of users
  - The type of content available to users (adult or restricted material, business or social content)
  - The modality of content available to users (image, text, audio, video)
  - How an audience consumes content (whether content is live-streamed, curated for specific communities/audiences, or sent using direct messaging)
  - The platform features or functionality (whether registration is required before content can be viewed, features such as downvoting, or the ability to live-stream content. Section 6 of this report also captures VSP features that enable the occurrence of specific harms)
  - Whether 'permanent' or ephemeral content is available to users.
- The scale of a service:
  - The number of users
  - The revenue generated by the service
  - The geographical reach or presence of the service
  - The number of employees
  - The average time spent per user per day on the service
  - The volume of video content hosted on the service
  - The volume of video content consumed by users of the service
  - The number of average monthly active users of the service.

Purpose

Consideration of the proportionality of regulatory measures is key for several reasons:

- The VSPs ecosystem is vast and diverse, as demonstrated by the many different factors relating to nature and scale captured above. Meta reported 304 million daily active Facebook users (not including Instagram or WhatsApp) in Europe from October to December 2022.[553] In comparison, the platform Yubo reports having 60 million total users globally, considerably less than Facebook.[554]
- If regulations are not proportionate to the scale of a service, there is a risk that obligations can become overburdensome and costly for smaller, newer platforms with more limited capacity and resources. Ofcom noted in a report summarising its first year of VSP regulation – that for some smaller platforms "a lighter touch analysis would be more appropriate and proportionate for these VSPs, especially in light of their more limited resources."[555]

- As part of the development of the Online Safety Bill in the UK, the Department for Culture, Media and Sports (DCMS) commissioned Impact Assessment research into the costs associated with a number of options for obligations regarding the Bill. It highlighted and explored transition costs (covering familiarisation with legislation, ensuring user reporting mechanisms are in place, and updating terms of service) and compliance costs (concerning conducting risk assessments, undertaking additional content moderation, age assurance technologies, transparency reporting, and reporting of online CSEA).[556]
- Regulation could also make a country a less attractive place to establish and run a digital business. DCMS highlighted this risk as part of its Impact Assessment of the preferred options for the UK's Online Safety Bill. A clear, proportionate, and effective regulatory framework was identified as part of the mitigation.[558]

7.3.1 Research summary

Online safety is a new area of regulation. Therefore, it is beneficial to explore other regulatory areas to understand if there are similar concepts or approaches. The concept of proportionality has featured in banking services regulation and food safety regulation. These approaches are summarised below.

The concept of proportionality within banking services regulation

- Proportionality is a key objective of banking services regulation. As part of European regulation, proportionality has been enshrined in legislation as a key objective for banking supervision. For European banking supervision, proportionality means "adapting the nature and intensity of supervision to specifics of the bank – its risk profile, its business model and its size".[559]
- The Chairman of the Financial Stability Institute noted in 2018 that the concept of proportionality "stems from the need to keep the level of public intervention – in the form of rules, restrictions or sanctions – appropriate to what is actually needed to achieve the desired social objectives".[560]
- Ensuring regulation is proportional in order to balance the reduction of risk and cost burdens for smaller organisations is acknowledged in the banking services sector. The European Central Bank (EBC) noted that as part of assessments into financial stability, the implications of proportionality for local competition need to be considered. In a statement from the Vice-Chair of the Supervisory Board of the ECB in 2017, the need to balance the costs of regulation and supervision (and that smaller banks face greater difficulties in complying with complex regulation), and financial stability risks, is highlighted. However, the Vice-Chair emphasised: "In my opinion proportionality means simpler rules for small banks. But it does not mean that the rules should be generally less stringent, or that banks can hold less capital or liquidity."[561]

The concept of proportionality within food safety regulation

- Articles 5 to 10 of the European General Food Law Regulation state that in matters of food safety, it is key to provide a systematic methodology to ensure "effective, proportionate and targeted measures or other actions to protect health."[562] Proportionate measures should be adopted to achieve a high level of health protection, while also accounting for the diversity in food supply and enabling an effective functioning market.

Varying levels of risk depending on the nature and scale of VSPs

- As highlighted throughout this report, certain VSP features can enable specific harms online. The presence of these features can increase the level of risk, particularly when considering the nature of the platform and the type of users it is aimed at.
- The scale of a platform may increase the level of risk due to the sheer number of users and the services' reach. For example, billions of posts are uploaded to Facebook, and more than two billion people use Instagram or Facebook daily.[563] Comparatively, Facebook and Instagram remove around 35,000 posts relating to self-harm and suicide every day, but missing even one percent of harmful posts leaves the risk of thousands remaining.[564] However, larger platforms may also have more resources and investment at their disposal to implement safety and mitigation measures.

- The level of risk on smaller platforms should not also be overlooked. A recent online safety report into post-digital intimacies from the University of Coventry noted, "...the focus on large technology companies such as Meta (i.e., Facebook, Instagram) TikTok, and Twitter means that communities of hate will still be able to exist in smaller, dedicated online spaces, and will continue to approach to precarious groups and at-risk individuals."[565]

Considerations for regulation

Cost of implementing provisions: It is important to consider the cost of any provisions that services will be obliged to implement, and whether this is proportionate to the nature and scale of the service. These costs will be in addition to any levies that are imposed on the service provider by the regulator. Taken together, the costs may result in providers deciding to cease operations or re-locate to another jurisdiction. The UK's Department for Culture, Media and Sport has published an Impact Assessment of the potential costs' businesses may occur in order to comply with the UK's proposed Online Safety Bill, which has been calculated as £250.6 million.[566]

Need for regulation to reflect the diversity of services in scope: The variation in nature and scale of platforms and the risks they pose requires regulation to not subscribe to a uniform approach, which fails to maintain proportionality.[567]

Proportionality of reporting obligations: It is important to consider how to prevent overburdensome reporting and ensure data collection or requests from regulators are straightforward and streamlined.

Burden of new regulation, specifically on small and medium-sized enterprises (SMEs): SMEs are recognised as an important, active part of any thriving economy. However, due to their size and often more limited resource, compliance with regulation can be challenging and navigation of the regulatory landscape burdensome. Consideration should be given to the potential impact of any provisions included with a code on SMEs, including the consideration of their particularities and the proportionality of the impact. The Organisation for Economic Co-operation and Development (OECD) noted: "SMEs face significant uncertainty when operating; consequently, when it comes to new regulation, some small business may end up bearing the costs without surviving long enough to enjoy its intended long-term benefits."

7.3.2 Potential areas for information-gathering

Table 28 below provides a summary of the types of information that might be sought and analysed to further explore the provision for proportionality. Some of this information is publicly available; other data would need to be requested from online service providers.

A Information that can be independently ascertained as a user of the service

B Information that can be ascertained from desk-based research using publicly available, third-party sources

C Information that would need to be requested from the specific service provider

D Information that can be obtained through the commissioning of new research

(b) The impact of automated decision-making on content moderation and content delivery processes	Information to assess	A	B	C	D
Type of users the service is aimed at	• Services' external branding and advertising campaigns • User surveys – existing user surveys could be reviewed or new research could be commissioned • Services' internal strategy documents			X	

Average age of users	• Marketing research could be commissioned specifically to determine the average age of users in Ireland	X		
Type of content available to users	• Landing page or timeline	X	X	X
Modality of content available to users	• Landing page or timeline			X
How audiences consume content	• User surveys – existing user surveys could be reviewed, or new research could be commissioned			X
Platform features or functionality	• Landing page or timeline			X
Availability of ephemeral content	• Landing page or timeline • Terms of service			X
Number of users	• Services' internal reporting and analysis			
Revenue generated	• Shareholder reports and information			
Geographical presence	• Shareholder reports and information • Services' website			
Number of employees	• Registration with Company Records Office or other regulatory bodies			
Average time spent per user on the service	• User surveys – existing user surveys could be reviewed, or new research could be commissioned • Services' internal reporting and analysis			
Volume of video content hosted	• Services' internal reporting and analysis			
Volume of video content consumed	• Services' internal reporting and analysis			
Average monthly active users	• Shareholder reports and information			

Table 28: Possible areas for information gathering on the need for provisions to be proportionate having regard to the nature and scale of the service

7.4 Availability, risk of exposure and risk of harm from harmful online content

This section provides a summary of available evidence of the levels of:

- availability of harmful online content on designated online services;
- risk of exposure to harmful online content when using designated online services; and
- risk of harm, and in particular harm to children, from the availability of harmful online content or exposure to it.

Interpretations and examples of key terms are provided in Table 29 below.

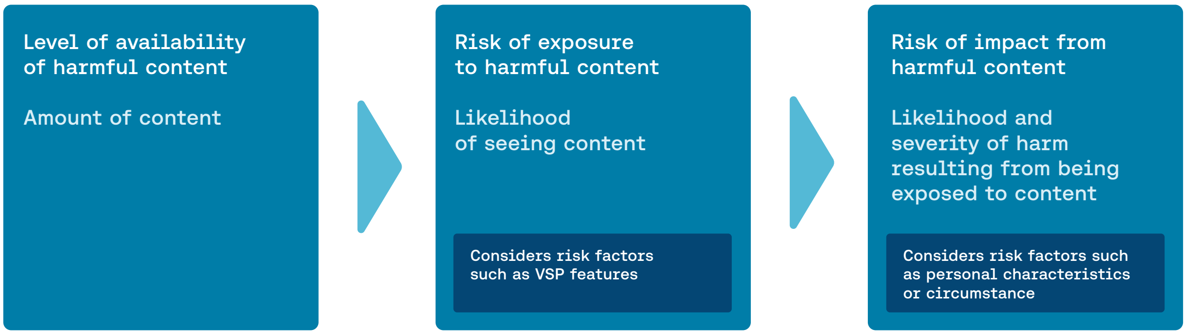
The relationship between the availability of harmful content, the risk of exposure, and the risk of harm is complex.[570] Individual users can experience varying levels of impact based on exposure to the same content; therefore, it is important to consider the measurement of these three aspects together to account for their complex relationship. Figure 5 demonstrates the interrelation and influencing risk factors considered within this report.

Purpose

Understanding the relationship between the availability of harmful content, risk of exposure,

Key term	Interpretation	Example relating to content promoting self-harm or suicide
Availability	The amount of online content in existence that is a potential cause of harm	A measure of the number of posts glorifying ‘depression’ or number of posts tagged with known hashtags related to promotion of self-harm or suicide behaviours
Risk of exposure	The likelihood of a user encountering harmful online content (where likelihood is affected by risk factors associated with VSP features and user behaviour)	The design of a VSP’s content delivery algorithm trained to deliver increasingly extreme content based on a user’s past behaviour
Risk of harm	The likelihood of a user being negatively impacted by experiencing online content (where likelihood is affected by risk factors associated with individual personal characteristics or circumstances)	The demographics of a VSP’s user base, or collective past experiences of those users within sub-communities (for example, users with a history of mental health issues)

Table 29: Terminology



Level of availability of harmful content

Amount of content

Risk of exposure to harmful content

Likelihood of seeing content

Considers risk factors such as VSP features

Risk of impact from harmful content

Likelihood and severity of harm resulting from being exposed to content

Considers risk factors such as personal characteristics or circumstance

Figure 5: The relationship between availability, exposure, and harm caused

and risk of harm is key to the identification of opportunities both to prevent harm, and to detect and address it when it does occur.

In an ideal scenario, regulation is informed by an understanding of the levels of availability, risk of exposure and risk of harm from harmful content, but also contributes to improve it. For example, this could be through the introduction of requirements such as risk assessments and mandatory reporting, which require providers to continually analyse platform risks, and report on their efforts to address them.

7.4.1 Research summary

Ofcom have commissioned research to develop a framework that maps the journey from harm exposure to impact, using the terminology of ‘hazards’, ‘risk factors’, and ‘harm’.[571] A summary of key findings demonstrates the complexities involved:

- There are multiple routes to experiencing harm because of online exposure. An isolated experience of a hazard (for example, a fraudulent advert) can lead to immediate harm or delayed harm. Alternatively, cumulative exposure to hazards over time can cause harm, this could either be through passive engagement (such as repeated exposure to a certain body type or content that makes certain unrealistic body shapes aspirational), or active engagement, (such as engaging in pro-anorexia communities online).
- Although harms resulting from an isolated incident that has an immediately harmful impact were the most recognisable to users, the research identified that the cumulative exposure of hazards resulted in the most severe harm experienced by respondents.

Another complexity in measuring online harm is the presence of risk and protective factors. Risk factors increase the probability of exposure occurring, or the probability and/or severity of harm occurring. Protective factors do the opposite. In general, risk factors associated with the risk of exposure will be determined by VSP features (see Section 6), response measures (see Section 7) and user behaviour. An example of user behaviour is children using adult content sites where there is a known risk of encountering CSAM content. Despite not being the child’s fault, the combination of the lack of age verification and the child seeking the content out creates the exposure risk.

Risk factors associated with the risk of harm will be determined by a user’s personal characteristics or circumstance, these might include protected characteristics (see Section 5.7), factors such as whether a child is in foster care, or protective factors – such as effective parental oversight.

The response to the issue of CSAM online is mature relative to the response to other harms, in large part due to the heinous nature of attendant crimes that have impelled global actors from all sectors to galvanise action to combat CSAM in recent decades. This is evident by the advanced understanding of the availability of child abuse content online, and of the risk of exposure and harm. Figure 6 below outlines the complexity of each of these matters and implications for how the issue of CSA online is tackled.

Levels of availability

The availability of CSAM is better understood than that of other types of harmful or illegal content, due to reporting mechanisms and automated detection. Key indicators include:

- The volume of reports of CSAM made to:
  - The National Centre for Missing and Exploited Children (NCMEC). Most of the largest global platforms report detected abuse imagery to NCMEC. NCMEC also takes reports from members of the public, although these account for a low proportion of overall reports.[572]
  - INHOPE, a global network of 52 hotlines, based in 48 different countries worldwide, for reporting CSA imagery. In 2022, INHOPE processed 587,852 reports of potential child abuse cases.[573]
  - The Internet Watch Foundation, a UK-based charity that also takes reports of CSAM, and undertakes some proactive work to detect such content online. In 2022, it assessed 375,230 reports of CSAM.[574]
- The volume of material proactively detected. Project Arachnid, a tool owned by the Canadian Centre for Child Protection (C3P), crawls the clear and dark web, and issues removal notices to electronic service providers when it detects CSA imagery. Since launching in 2017, it has detected more than 42.7 million such images.[575]
- Platform transparency reports. Included in such reports for most of the larger platforms is data regarding the number of identified cases of online CSA. For example, between 1 January 2022 and 31 December 2022, Meta actioned 6.6 million pieces of media constituting CSAM in threads including an EU-based user.[576]

Although varied and trusted, the above indicators do still have limitations. Broadly speaking, tools and technologies to detect 'known' CSAM (imagery which has already been reported) are more mature and more widely adopted. Additionally, uptake of tools to detect CSAM in video format is also lower – and video accounts for an increasing proportion of such material. It is therefore likely that, despite the relative maturity of reporting and detection of CSAM, actual availability is higher than indicators would suggest.[577]

Risk of exposure

Risk of exposure can be created, reduced, or removed by VSP features and subsequently realised through user behaviours. A broad range of VSP features can impact the risk of exposure to CSAM. Examples include:

- Tools to detect and block CSAM before it is possible for a user to see it (such as content classifiers and hash-based matching – see Section 5.10) can remove or reduce the risk of exposure. Such tools can be deployed at device, network, or platform level.
- Seamless sharing can significantly increase the risk of exposure to CSAM when the above tools are not implemented. As an illustrative example, between October and November 2020, Facebook found that copies of just six videos were responsible for more than half of the child exploitative content reported in that time period.[578]

User behaviour is a secondary factor that can impact risk of exposure. Recognition of this is implicit in the provisions of the UK Online Safety Bill, which requires service providers to factor "the different ways the service is likely to be used and in turn... the risk of harm that might be suffered by individuals because of these methods of use."[579]

An illustrative example relevant for the case of online CSA content are adult sites. CSAM has been detected on such sites,[580] and it is well documented that minors access them. This demonstrates how the lack of VSP safety features (such as age verification) permits user behaviour (young people seeking adult content), which in turn creates the exposure risk.

Risk of harm

Risk of harm is much more subjective. It depends on a range of circumstantial and personal factors.

Circumstantial factors that can affect whether, and to what extent, harm is experienced as a result of exposure to harmful and illegal content include:

- An individual's immediate family circle and support: Parents and caregivers can play an important role in keeping children safe online via use of parental control mechanisms and establishing limits on screen-time and internet usage. Effective parental oversight is an example of a protective factor that can reduce the risk of harm. Support provided by parents and guardians can also reduce the severity of harm.

- Community and culture: Stigmas or trends indirectly fuel higher risk behaviours. For example, there is some evidence to suggest that LGBTQ+ children are more likely to feel comfortable connecting online with individuals aged 21 and over.[581] It is possible that societal attitudes mean such children feel less able to connect with others in the real world and therefore turn to the internet to establish relationships, which can carry risk.

- Country: In China, online pornography is banned, which means that it can only be accessed by circumventing centrally implemented measures to block citizens' access.[582] Country-level approaches can also impact an individual's access to support mechanisms. As an example, in some countries, boys are ineligible for protective measures and recovery programmes for survivors of sexual abuse.[583]

- Education and awareness: These protective factors can reduce the risk or potential severity of harm by increasing consciousness of harms, helping users to proactively manage their risks of exposure and seek support.

- Internet regulation: New laws in the UK, Australia and Europe require service providers to detected and report CSAM. As new internet regulators become established, there should be more evidence of how such provisions reduce the risk of harm.

Personal factors relevant to the risk (and severity) of harm include:

- Disability: Studies have shown, for example, that deaf children are more likely to spend extensive time online, possibly due to limitations moving freely in the physical world and to avoid stigma associated with disabilities.[584]

- Age: There is some evidence to suggest that younger children are more severely impacted by exposure to certain content, for example online pornography.[585]

- Socioeconomic status/poverty: This can impact an individual's access to support mechanisms should they come to harm. While poverty can worsen harm through a lack of recourse to trauma-informed support, for example, access to effective support can reduce the severity of harm and is therefore an important protective factor.

Figure 6: CSAM: Measuring availability, risk of exposure, and risk of harm

Considerations for regulation

Having an accurate understanding of the availability of harmful content, the risk of exposure, and the risk of harm is key to effective regulation. It should inform the provisions included in mandatory Codes, and any updates to rules that may be required to address emerging risks/harms. The various information sources that can contribute to developing the requisite understanding are detailed in Table 30 (Figure 6 above includes example metrics from some categories of source).

It is also worth noting that many of the above sources have limitations. Broadly speaking, these can be grouped into three areas:[586]

1. Most potential measures/sources were not designed with the intention of being used as a measure of online harm, for example:
   a. Transparency reports provide information about the action taken by platforms on content that violates their community guidelines. The development of these guidelines and reports were likely driven by a range of factors, including protecting users, publicity, and politically or socially relevant issues based on the locations they operate in. While the data these reports provide may be a useful indicator about some harm areas, their primary purpose is not to provide an objective assessment of harm.
2. Technical limitations of research methods employed, for example:
   a. Reliance on self-reported internet data usage, which is limited and can be problematic due to:
      i. Low recall accuracy and subjective interpretation of questions
      ii. Social desirability bias
      iii. Users' inability to report on harms that they do not recognise
   b. Measures tend to focus on short-term impacts
   c. Automated tools are not able to understand context.
3. Lack of consistent definitions and granularity, for example:
   a. Inconsistent definitions of online harms used in different measures
   b. A lack of granularity or detail about what hazards people are exposed to
   c. Few measures account for how often people are exposed to hazards
   d. There is limited detail around which users experience harm (such as children), and on which platforms (including encrypted spaces).
   e. Most measures ignore positive outcomes, including those related to fundamental rights discussed in Section 8.5.

Under-reporting is an important factor that can undermine the value of user reports,[587] as it affects the accuracy of estimating the size and scale of the problem. This is discussed in most of the harm profiles in Section 0. Under-reporting is an issue globally, including in Ireland. A 2021 survey of country users revealed that only eight percent of 9–17 year-old internet-using Irish children reported harmful content to the website or service provider.[588] Under-reporting is further compounded by the lack of independent analysis of current transparency reports (see Section 7.1).

There is a gap in our understanding of the relevant matters due to limitations with available information sources. For example, we do not understand the impact of exposure frequency on harm caused or whether seeing extreme content once is less harmful overall than being frequently exposed to less severe content. A more detailed example of this is provided in Section 5.4 (eating disorders). Additionally, most sources provide a snapshot and do not reveal the impact or harm caused over time.

7.4.2 Potential areas for information-gathering

As the regulator responsible for determining the requirements companies must fulfil in order to keep users safe, it is imperative that An Coimisiún has access to all relevant information that can provide an insight into matters (d), (e), and (f). For this reason, it should consider the areas for evidence-gathering detailed in Table 31 below.

A Information that can be independently ascertained as a user of the service

B Information that can be ascertained from desk-based research using publicly available, third-party sources

C Information that would need to be requested from the specific service provider

D Information that can be obtained through the commissioning of new research

Evidence	Suggested information	A	B	C	D
Number of pieces of harmful content removed from platform	• Reports published by social media platforms containing information about the platforms' guidelines and policies, and data on how these are enforced		X	X	
Number of potentially harmful pieces of content on a platform	• Using automated tools (AI, machine learning) to identify content or behaviour • Academics often use automated tools when analysing samples of social media data (for example, a dataset of tweet), to detect the prevalence of potentially harmful content. These are referred to as 'measurement studies'		X		
Evidence of the displacement of harm' manifestation in 'hidden' online spaces (for example, within end-to-end encrypted environments).	• Evidence from law enforcement pertaining to potentially adverse effects of 'surface web' regulation, to understand the risk of harm displacement into other domains				
Number of exposures to harmful content	• Reports published by social media platforms containing information about the platforms' guidelines and policies, and data on how these are enforced			X	
Number of [x] users	• Where [x] relates to a protected characteristic			X	
Longitudinal evidence, of the short- and long-term impact of online harm	• Regular reviews of research, data, or provided information		X	X	
Harm impact research that analyses the experience of harm for groups based on protected characteristics, factoring both positive and negative effects	• Behavioural and attitudinal studies that include detailed analysis of harm impact over a long period of time		X		X
Harm reporting data from third parties and law enforcement	• For example, the number of NCMEC referrals or web pages actioned by the Internet Watch Foundation (IWF) for hosting illegal content		X		

Table 31: Example areas for information-gathering regarding the relevant matters

7.5 User rights

Consideration of the rights of users of designated services must be based on an understanding of the following key charters, laws, and guidelines:

1. International and national legislation that enshrines human rights:
   - Charter of Fundamental rights of the European Union (EU),[589] which sets out the rights enjoyed by the people of the European Union, promoting human rights within the EU.[590]
   - These fundamental rights are also reflected in the Constitution of Ireland.[591]
2. International and national legislation that enshrines digital rights:
   - The Digital Services Act (DSA) outlines the regulations that apply to online services acting as intermediaries between users and goods, services, or content – granting users protections and fundamental rights online.[592]
3. Principles or guides that set out the rights of users online:
   - The Human Rights for Internet Users by the Council of Europe[593] is a guide produced by the Committee of Ministers of the Council of Europe to ensure that the fundamental rights guaranteed by the European Convention on Human Rights are applied to internet users through the principles set out in this framework.

Although VSPs are also covered by Irish Consumer Law,[594] the protections that consumer law grants users mainly cover the provisions of these services in accordance with Terms and Conditions developed by providers themselves.

Definition

The rights with particular relevance for online safety and An Coimisiún's regulatory responsibilities are:

- Human dignity, and physical and mental integrity
- Liberty and security of person
- Privacy
- Protection of personal data
- Freedom of expression
- Academic freedom
- Protection of children
- Access to the internet
- Assembly, association, and participation
- 'Rights to remedy'.

Detailed descriptions of these rights are included in Table 34 (annexed).

Purpose

Consideration of user rights is important to ensure continuity of protections across the 'real' and 'online' worlds. This is key due to the increasing convergence of these realms, and the fact that people transact an increasing proportion of daily activities online. A failure to consider the rights of users of designated services can result broad breaches of fundamental human rights, causing both online and real-world harms.

However, while online safety legislation is a route to embedding respect for user rights, it can also inadvertently contribute to rights being infringed, for example if safety provisions unduly restrict freedom of expression. An article in the UN Chronicle outlined examples where user rights had not been duly incorporated into online safety regulation, leading to potential criminalisation of free expression and censorship.[595] The need for transparency reporting to support the upholding of user rights was subsequently highlighted through the Online Harms White Paper consultation (2019), resulting in it becoming a key recommendation of the UK government report on transparency reporting in relation to online harms (2020).[596] For the issue of online harms, the challenge of protecting and upholding user rights is complex.

7 .5.1 Research summary

Regulation that may impact certain rights is not a novel concept. Governments regulate goods with negative and positive externalities in our every-day lives, for example, putting in place age restrictions for smoking and alcohol. The complication with online safety and user rights is the scale at which information is shared, and therefore the possible magnitude of user right infringement. Online safety regulation needs to delicately balance safety and security in the digital world with freedom of expression, privacy, and data protection – making this a very complex topic.

The harms described in part one of this report demonstrate the necessity for online safety regulation, and further, that the impact of some of the harms can directly contradict people's rights to human dignity, physical and mental integrity, liberty, security of person, privacy, and protection of personal data. However, the oversight and regulation from governments required to combat these harms can also conflict with the right to freedom of expression, privacy online, and data protection. Pressure for further consideration of privacy and freedom of expression in online safety legislation is common:

- When the Online Safety and Media Regulation Bill was being debated in 2021, the Irish Council for Civil Liberties expressed concern about the impact to freedom of expression and other rights if non-illegal speech becomes restricted online.[597]
- In 2022, six rights groups addressed a letter to Irene Khan (UN Special Rapporteur) asserting that the UK government's Online Safety Bill undermines "international human rights principles".[598]

Discussion of impact

Freedom of expression can stand in conflict with online safety, as content or account removal for safety reasons may be construed as directly infringing an individual's right to freedom of expression. Reconciling the two is difficult, as it is known that the ability for users to create content and upload it can lead directly to harm (see part one of this report). This trade-off is also acknowledged in Ofcom's statement after the Buffalo, New York terrorist attacks.[599] In recent years, several high-profile cases have raised the question of whether full freedom of expression and content moderation can co-exist. For example, Donald Trump's removal from Twitter after the capitol riots[600] drew direct criticism from then-German Chancellor Angela Merkel, who suggested that this was 'problematic' due to the right to 'freedom of opinion'.[601]

Similar complexity surrounds the topic of privacy and online safety/security. Encryption and anonymity are common measures to ensure privacy in the digital world. Encryption is the protection of content or data through mathematical models that require 'keys' to access them.[602] However, this removes the ability to monitor content and can therefore enable the propagation of harm through, for example, the sharing of CSA or terrorist material. The conflict of the two is demonstrated in several recent high-profile cases.

- CSAM: In 2021 Apple planned to scan pictures in the iCloud for CSAM but halted this after push-back from rights groups on privacy concerns.[603]
- Terrorist content: In 2019, Apple opposed the FBI's request to de-crypt its operating system after the San Bernardino terrorist attack, stating that this could undermine citizen's freedoms and liberties.[604]

Messaging services WhatsApp and Signal have previously made public statements opposing potential requirements for the removal of end-to-end encryption.[605, 606] Proponents of end-to-end encryption have argued that it is critical, particularly in regions of the globe where individuals and groups are at risk of persecution, for example by corrupt regimes.

Considerations for regulation

There are two key elements that should guide the regulation of VSPs in terms of user rights considerations. Firstly, all applicable human and user rights need to be considered, as they may sometimes oppose each other in the digital context. These need to be weighed up against each other. As an example, if someone's right to privacy enables CSAM to be propagated, this directly infringes upon the child's right to protection through the harm it causes the child. Here, we are not just dealing with the right to privacy, but also the right to security, children's protection, protection of personal data, and mental integrity – all complex rights that need to be balanced. Guides such as the Human Rights for Internet Users by the Council of Europe include significant caveats within its freedom of expression and privacy sections, such as the exclusions of expressions inciting discrimination, hatred or violence, and the possibility of intervention in privacy due to law enforcement.[607] This may provide a useful guiderail when balancing user rights in new legislation. Secondly, it is important to weight the potential prevalence and severity of harm against the degree of intrusion: something like CSA (the effects of which are proven to be severe for both individuals and society) may justify more intrusive countermeasures than a harm like cyber bullying, for example.

The use of AI for content moderation complicates matters such as freedom of expression further, and these methods need to be carefully examined when regulating VSPs. As discussed in Section 8.2, automated content moderation can lead to certain groups being marginalised and having their freedom of expressions curbed more than others. Investigations into how VSPs can uphold human rights (and thereby user rights) in content moderation protocols are also required, as certain VSPs have recently been accused of hindering the documentation of human rights violations.[608]

The risks and consequences of not considering user rights in online safety legislation can be grave. For example, a North African UN state implemented 'cyber-crime' legislation ordering internet service providers to keep users' data (such as phone calls, text messages, and browsing history), and to allow law enforcement access without safeguards. It is alleged that this opened the door to state censorship and online surveillance, subsequently reducing civil liberties, freedom of expression, and the privacy of residents and citizens.[609]

On the other hand, the failure to introduce online safety legislation essentially leaves platforms to determine what is 'right' and 'wrong'; what constitutes privacy, freedom of expression, and sufficient data protection – based on their commercial agendas.

7.5.2 Potential areas for information-gathering

On the surface, an analysis of VSPs' Terms and Conditions or community standards can be used for evidence-gathering, as these outline the principles that VSPs apply to moderate their platforms. Surveys can also be leveraged to understand whether users perceive their rights to be respected and upheld by providers.

For a more detailed analysis, the UK government report on transparency reporting in relation to online harms (2020)[610] recommended transparency reporting as a tool to assess VSP compliance with user rights. It stated that such reports should ideally outline the "measures and safeguards that are in place to uphold and protect user rights" and their effectiveness. Part of this will include the decision rationale to remove content and/or accounts, the quality assurance of moderation decisions, and the process of appeal. It is important to understand how companies balance content moderation and removal with the compliance with user's freedom of expression, as there is a risk that transparency reporting of harms could lead to over-removal of content. Therefore, transparency reports should not only focus on the quantity of harmful content. For further detail, refer to Section 8.1. It is worth noting that certain VSPs already produce transparency reports , as detailed in that section. Lastly, there are several organisations that An Coimisiún may want to collaborate with to understand VSPs' compliance better, such as the Competition and Consumer Protection Commission (CCPC). This would help explain any general infringements of consumer law by VSPs. Collaboration with the Advertising Standards Authority for Ireland could identify any marketing infringements.

A	Information that can be independently ascertained as a user of the service
B	Information that can be ascertained from desk-based research using publicly available, third-party sources
C	Information that would need to be requested from the specific service provider
D	Information that can be obtained through the commissioning of new research

(g): The rights of users of designated online services	Suggested information	A	B	C	D
Principles of content moderation and removal (rationale for removal)	• Terms and Conditions/Community Standards • Transparency reports		X		
Measures and safeguards in place to uphold and protect user rights, including their efficacy	• Transparency reports • Direct information requests		X	X	
Information on the moderation process and quality assurance of such moderation decisions, including the number of content removals	• Transparency reports • Direct information requests		X	X	
Measure user awareness of the appeal processes, number of appeals and error rates of moderation processes	• Transparency reports • Direct information requests		X	X	
Information on the use of algorithms and automated processes in content moderations	• Transparency reports • Direct information requests		X	X	
Perceptions and user sentiments	• Surveys		X	X	X

Table 32: Possible areas for information gathering for the compliance with the rights of users of VSPs[611]

7.6
Service provider rights

This section explores the rights of providers of VSPs for consideration within online safety regulation. Similarly to user rights, provider rights are enshrined in various laws:

1. EU competition legislation
   - Treaty on the Functioning of the European Union (TFEU): One of two primary treaties forming the constitution of the European Union.[612]
   - Digital Markets Act: A European Union regulation aimed at defining and regulating 'gatekeepers' within the digital market.[613]
2. Irish competition legislation
   - Competition Act 2002 (among other relevant/ adjacent acts and amendments): Irish legislation that ensures fair competition in Irish markets.[614]
3. Constitution of Ireland
4. EU human rights legislation
   - Charter of Fundamental rights of the European Union[615] setting out the rights enjoyed by the peoples of the EU, to promote human rights within the EU.[616]

Definition

Notable rights to consider for VSPs providers are the right to freedom to conduct a business, fair competition, and protection of intellectual property.

Purpose

Akin to the importance of users' rights, the consideration of provider rights is imperative when designing online regulation for Ireland. Insufficient consideration of existing legislation and agreements could lead to future incompatibility of the Online Safety Code and further unintended negative consequences. Similarly to proportionality (covered in Section 8.3), these impacts may indirectly distort fair competition or the right to conduct business.

7.6.1 Research summary

In the effort to combat online harms, future regulation of VSPs will likely require transparency from providers and prescriptive regulations. Despite reduction in online harms being the main objective, there are many other impacts that such regulation will have. The evaluation of these impacts is key to understanding the consequences of regulation, and therefore the compatibility with provider rights – such as fair competition, the right to conduct a business, and protection of intellectual property.

Considerations for Regulation

Effective regulation of VSPs will require transparency from providers on their practices and principles. For example, sharing moderation principles, processes, and appeals statistics is useful in ensuring user rights are upheld, as discussed in Section 8.5. However, such information may be core to the business models of certain VSPs and deliberately kept confidential to reduce the risk of competitors gaining privileged information. Therefore, the (potential) conflict arises when the transparency required leads to the (often inadvertent) disproportionate favouring or disadvantaging of one or a few VSPs, as this infringes on providers' rights to fair competition and protection of intellectual property. The ask on providers for more transparency needs to be designed so that it does not distort 'fair' economic market conditions or infringe on IP law.

Further, regulation will likely be prescriptive in its asks, and while uniformity of regulation is useful for the regulation of harms (and its evaluation), firms also have the 'freedom to conduct a business' in line with the Charter of Fundamental rights of the European Union.[617] This will come with the expectation and right to a certain level of autonomy, bounded within EU and national laws. As an example, while competition law regulates that VSPs need to provide digital services to users as described within their Terms and Conditions, it does not mandate what these Terms and Conditions are. The right for VSPs to design their own Terms and Conditions is also implied through the guide to Human Rights for Internet, exemplified in the prior mention of Twitter's removal of Donald Trump after the violation of the platform's Terms and Conditions.[618] Regulation of VSPs will impact their autonomy and this must be carefully balanced, weighing consideration of possible disbenefits to companies against likely benefits to users.

The regulation of VSPs without careful consideration of the providers' rights (over-regulation) could lead to a removal of platform autonomy or impede company rights to fair competition, and may result in companies relocating to evade such restrictions. In addition to potentially negative economic consequences, this could result more harm being caused to Irish users if platforms relocate to unregulated jurisdictions and services can still be accessed.

7 .6.2 Potential areas for information-gathering

Based on the above analysis of provider rights (freedom to conduct business, fair competition, and IP protection), there are limited areas for evidence gathering.

A	Information that can be independently ascertained as a user of the service
B	Information that can be ascertained from desk-based research using publicly available, third-party sources
C	Information that would need to be requested from the specific service provider
D	Information that can be obtained through the commissioning of new research

(g): The rights of providers of designated online services	Suggested information	A	B	C	D
Collaboration with European authorities on current cases relating to company rights	• Examples of recent cases that have generated relevant rulings and precedents include the merger of WhatsApp and Facebook in 2014[619] (granted by the European Commission), and the Statement of Objection for Meta's conduct with advertisement tying in Facebook Marketplace.[620]		X		
Company responses to online regulation in other parts of the world	• Press releases regarding company operating decisions in response to legislation, such as WhatsApp's opposition to encryption removal and Signal's announcement that it may choose to no longer serve UK customers when the Online Safety Bill becomes law.[621, 622]		X	X	

Table 33: Possible areas for information gathering for the compliance with the rights of users of VSPs[623]

8

Further areas of potential interest

Research into online harms is comparatively new, and much of the literature on this topic is emerging and developing. The research undertaken as part of the drafting of this report has identified particular areas where the online harms literature is particularly nascent. An Coimisiún may wish to periodically review new research or literature that is published in these areas to consider if it is beneficial and relevant for any potential future iterations of the Online Safety Code, or to inform more general awareness of issues in relation to harmful online content.

New research or literature could be conducted by a range of organisations including, but not limited to, NGOs, other regulators, third-sector bodies, or online platforms. An Coimisiún may also wish to commission research itself on the topics listed below, in additional to examining future relevant material published by other organisations.

- Irish users' attitudes to and perceptions of VSPs (covering all online harms): An Coimisiún may wish to consider any future publicly available attitudinal research or surveys fielded specifically to the Irish population. Research in this space may provide additional context on the issue of online harms and VSP usage.
- Longitudinal studies on the impact of all online harms: Studies of this type have been limited to date in part due to the ethical considerations, costs, and difficulties associated with including 'control groups.' An Coimisiún may wish to consider evaluating findings from any future longitudinal studies that are published on the impact of harmful online content, as this research would aid understanding of risk and protective factors that increase or reduce online harm. Within this, future research regarding the impact on groups with specific risk factors and vulnerabilities (for example children with disabilities, or who identify as LGBTQ+) could warrant consideration.

- Prevalence research on all online harms: An Coimisiún may wish to consider examining future research that is published on the prevalence of all types of harmful content. Organisations may conduct this research using automated solutions, as well as self-reporting data, and platform transparency reports.
- Harms-specific research: At present, research on the impact of certain harms on people, and in particular children, is limited. In particular, there is comparatively less literature on harms related to 'gratuitous violence and its impact on children', suicide, disordered eating, and audio-visual commercials. An Coimisiún may therefore find it beneficial to consider any future research published on these topics, especially if large sample size surveys are used.
- Impact of future technologies: Future technologies, such as haptic suits and the metaverse, will likely have a range of implications for online safety. An Coimisiún may therefore wish to consider any emerging research and literature published in the future on these issues, and in particular research exploring the potential consequences for regulation or how potential new harms can be reduced or 'designed out'.

Annex A

Further information on user rights

User rights topic	Charter of Fundamental rights of the European Union[624]	Human Rights for Internet Users by the Council of Europe[625]	Digital Services Act[626]
Human dignity, and physical and mental integrity	Article 1: Human dignity is inviolable. It must be respected and protected. Article 3.1: Everyone has the right to respect for his or her physical and mental integrity.	–	The protection of human dignity is covered through better protection of victims of cyber violence, including the swift take down of non-consensual sharing of illegal content when reported.
Liberty and security of person	Article 6: Everyone has the right to liberty and security of person.	–	N/A
Privacy	Article 7: Everyone has the right to respect for his or her private and family life, home, and communications.	Users have the right to private and family life on the Internet which includes the protection of personal data and respect for the confidentiality of correspondence and communications. This also includes that: • Public authorities and private companies have an obligation to respect specific rules and procedures when they process personal data • Personal data should only be processed when laid down by law or when you have consented to it • Users must not be subjected to general surveillance or interception measures.	

Personal data	Article 8: 1. Everyone has the right to the protection of personal data concerning him or her. 2. Such data must be processed fairly for specified purposes and on the basis of the consent of the person concerned or some other legitimate basis laid down by law. Everyone has the right of access to data which has been collected concerning him or her, and the right to have it rectified. 3. Compliance with these rules shall be subject to control by an independent authority.	Please see the 'privacy' row on this.	Targeted advertising when using sensitive data such as sexual orientation, etc. is banned. Similarly, behavioural nudging to get users to use their services is also banned. Please also see the 'Freedom of expression' row.
Freedom of expression	Article 11: 1. Everyone has the right to freedom of expression. This right shall include freedom to hold opinions and to receive and impart information and ideas without interference by public authority and regardless of frontiers. 2. The freedom and pluralism of the media shall be respected.	Users have the right to express themselves freely online and access information, opinions, and expressions of others. This includes, creation, re-using and distributing content, as well as not disclosing your identity online.	Notices should be processed with respect to the freedom of expression and data protection, in a non-arbitrary and non-discriminatory manner.
Academic freedom	Article 13: The arts and scientific research shall be free of constraint. Academic freedom shall be respected.	–	–

Protection of children	Article 24: 1. Children shall have the right to such protection and care as is necessary for their well-being. 2. In all actions relating to children, whether taken by public authorities or private institutions, the child's best interests must be a primary consideration.	Children and young people are entitled to special protection and guidance when using the Internet, such as protection from interference with your physical, mental, and moral welfare, in particular regarding sexual exploitation and abuse on the Internet and other forms of cybercrime.	Minors have the right to be protected on platforms, including from targeted advertising. Please also see the 'Human dignity, physical and mental integrity' row.
Access to the internet	Article 36: The Union recognises and respects access to services of general economic interest as provided for in national laws and practices, in accordance with the Treaties, in order to promote the social and territorial cohesion of the Union.	Users should have affordable and non-discriminatory access to the Internet.	–
Assembly, association, and participation	–	You have the right to peacefully assemble and associate with others using the Internet.	–
Rights to remedy	–	You have the right to an effective remedy when your human rights and fundamental freedoms are restricted or violated.	Users have the right to compensation for damages or loss resulting from platform non-compliance.

Table 34: List of topics that should be covered under user rights and the sources that support these

Glossary

A

ACC
Audio-visual Commercial Communication.

Alt accounts
Having multiple accounts on the same service.

Anonymity
The absence of personally identifiable information.

AVMSD
Audio-Visual Media Services Directive.

C

CARI
Children at Risk in Ireland.

Child sexual abuse
The involvement of a child (anyone under 18) in sexual activity that they do not fully comprehend, are unable to give informed consent to, or for which the child is not developmentally prepared and cannot give consent.

Child sexual abuse material (CSAM)
Any visual or audio content of a sexual nature involving a person under 18 years old, whether real or not.

Child sexual exploitation
Abuse that involves any actual or attempted abuse of position of vulnerability, differential power, or trust.

Child sexual exploitation and abuse (CSEA) online
Abuse and exploitation partly or entirely facilitated by technology.

Clickbait
Online content intended to attract attention and encourage people to click on links to particular websites.

Communities
A collection of users who share similar interests in the content and content creators they engage with.

Community reporting/flagging
Reporting mechanisms allowing users to flag self-harm content for review.

Content classifiers
Algorithms or human moderators can review and restrict potentially harmful content.

Content delivery processes
Processes for the delivery of content (audio, visual, video media) to users of online services.

Content moderation

Processes for reviewing online user-generated content for compliance against policies on what is and what is not permitted to be shared.

Content recommendation

Algorithms used by VSPs to recommend content to users. Content with more interaction will be recommended to more users.

Cyber bullying

Bullying with the use of digital technologies. It can take place on social media, messaging platforms, gaming platforms, and mobile phones.

Cyber harassment

Harassment taking place online.

Cyber stalking

Stalking that takes place online.

D

Defined networks

Used by some services to allow users within a defined network (such as like those with a school email or geographical tag) to post anonymous messages.

Digital pruning

The act of selecting who users follow so they can manage what content they see.

Direct messaging

Enables users to engage in communication privately, including group chats.

Disappearing/transient content

Content that expires after a certain amount of time and encourages users to share in the moment.

Disinformation

Intentionally false or misleading information that can take many forms – from memes to low-quality clickbait.

Doxing

The practice of gathering and publishing personal or private information about someone on the internet.

E

Eating disorders

Often also referred to as ‘pro-anorexia’, ‘pro-ana’, ‘pro-bulimia’, or ‘pro-mia’; it is online content that promotes the harmful behaviour and mindset that forms part of eating disorders.

Encryption

The protection of content or data through mathematical models that require ‘keys’ to access them.

End-to-end encryption

A message is encrypted at the sender's end and decrypted on the receiver's end. The message remains encrypted at all points during transit, so even if someone intercepts it during transmission, they can't read its contents.

Endless scrolling feeds and autoplay

VSP design that allows for endless scrolling of content that plays automatically, loops upon ending, or automatically moves the user onto the next recommended video.

Engagement mechanisms

The ability to engage with other user’s content to provide feedback (for example, likes, comments, or shares).

F

Feedback mechanisms

A process that provides feedback to the user who reported inappropriate or harmful content.

Flaming

Posting personal insults and vulgar and angry words; an intense argument and very aggressive form of intimidation that can take place in online spaces, including VSPs.

G

Generative artificial intelligence (or generative AI)

A type of artificial intelligence system capable of generating text, images, or other media in response to prompts.

H

Haptics

The use of technology that stimulates the senses of touch and motion, especially to reproduce in remote operation or computer simulation the sensations that would be felt by a user interacting directly with physical objects.

I

Influencer

A user high in social standing who has the power to affect their followers' beliefs and purchasing decisions.

INHOPE

A global network of 52 hotlines, based in 48 different countries worldwide, for reporting child sexual abuse imagery.

Internet Watch Foundation

UK-based charity which that takes reports of child sexual abuse material, and undertakes some proactive work to detect such content online.

ISPCC

Irish Society for the Prevention of Cruelty to Children.

L

Live-streaming

Technology that lets you watch, create, and share videos in real time.

M

Manosphere

An online movement of anti-feminist websites and communities focused primarily on 'men's issues.'

Metaverse

A metaverse is a 3D, virtual space facilitated by technologies – including virtual reality (VR), augmented reality (AR), and Internet of Things (IoT) – that allows people to interact with each other.

N

NACOS

National Advisory Council for Online Safety.

NCMEC

National Centre for Missing and Exploited Children.

Neuromarketing

Advertising aimed at influencing users at a subconscious level.

O

Ofcom

The UK media and telecoms regulator.

Online sexual harassment

Unwanted sexual conduct that occurs on digital platforms.

OSMR

Online Safety and Media Regulation Act 2022.

P

Pseudonymised usernames

Popular on VSPs, where usernames do not reflect a user's legal name.

R

Randomised meets

Services that offer video chats between randomised and anonymised individuals for real-time interaction.

Recommendation systems

Used for item/content/network filtering based on user preferences and/or past behaviour.

Revenge porn

Non-consensual sharing of sexualised images online.

S

Santa Clara Principles

Three principles on transparency and accountability around internet platforms' use of user-generated content, launched by a group of human rights organisations, advocates, and academic experts.

Seamless sharing

Easily accessible lists or groups of contacts make the wide sharing of content seamless.

Social bots

Artificial accounts that emulate human communication.

T

Targeted advertisements

Advertisements that are specifically aimed at users based on their characteristics, preferences, or past engagements.

Transient/disappearing content

Content that expires after a certain amount of time and encourages users to share in the moment.

Trolling

When someone posts or comments online to deliberately upset others.

Twitter

Now known as X.

U

Upload filters

Mechanisms by which content is scanned either at the point of upload or prior to publication.

User generated content

Any form of content that has been posted by users on online platforms, tailored or crafted for a user's feed.

V

Visual editing

Image manipulation via filtering or editing by computer software. These edits are often undetectable.

VLOPs

Very Large Online Platforms (platforms or search engines that have more than 45 million users per month in the EU).

VSP

Video-sharing platforms.

Endnotes

Introduction

[1] NACOS: Report of a National Survey of Children, their Parents and Adults regarding Online Safety (2021), available at https://www.gov.ie/en/publication/1f19b-report-of-a-national-survey-of-children-their-parents-and-adults-regarding-online-safety/

2 | Scope

[2] AVMS Directive (EU) 2018/1808 (2018). Available at: https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32018L1808&rid=9

5 | Online harms profiles

5.1 | Online harm prevalence in Ireland

[3] NACOS: Report of a National Survey of Children, their Parents and Adults regarding Online Safety (2021)

[4] Barnardos, Cyber Bullying Report (2022) Available at: https://www.barnardos.ie/barnardos-launches-cyberbullying-report/

[5] NACOS: Report of a National Survey of Children, their Parents and Adults regarding Online Safety (2021)

[6] Cybersafe Kids, Academic Year in Review (2021-2022). Available at: https://www.cybersafekids.ie/wp-content/uploads/2022/09/CSK_YearInReview_2021-2022_FINAL.pdf

[7] Barnardos, Cyber Bullying Report (2022)

[8] NACOS: Report of a National Survey of Children, their Parents and Adults regarding Online Safety (2021)

[9] NACOS: Report of a National Survey of Children, their Parents and Adults regarding Online Safety (2021)

[10] NACOS: Report of a National Survey of Children, their Parents and Adults regarding Online Safety (2021)

[11] The Irish Examiner, 'Shocking' link between viewing of porn and child-on-child abuse' (2023). Available at: https://www.irishexaminer.com/news/arid-41139007.html

[12] NACOS: Report of a National Survey of Children, their Parents and Adults regarding Online Safety (2021)

[13] Women's Aid, One in five (2020). Available at: https://www.womensaid.ie/app/uploads/2023/04/One-in-Five-Youn-Women-Report-2020.pdf

[14] FRA (EU agency for fundamental rights), Fundamental rights survey (2019). Available at: https://fra.europa.eu/en/data-and-maps/2021/frs?mdq1=theme&mdq2=982

[15] Western People, Hotline ie removed over 14,000 pieces of online child sexual abuse material in 2021 (2022). Available at: https://westernpeople.ie/2022/12/21/hotline-ie-removed-over-14000-pieces-of-online-child-sexual-abuse-material-in-2021/

[16] Competition and Consumer Protection Commission, Online behaviour: Influencer marketing (2022) Available at: https://www.ccpc.ie/business/wp-content/uploads/sites/3/2022/12/2022.12.12-172837-CCPC-Influencer-marketing-report.pdf

5.2 | Use of video-sharing platforms (VSPs)

[17] Cybersafe Kids, Academic Year in Review (2021-2022).

[18] NACOS: Report of a National Survey of Children, their Parents and Adults regarding Online Safety (2021)

[19] Cybersafe Kids, Academic year in review 2021 – 2022. (2022)

[20] OFCOM, Children and parents: media use and attitudes report 2022. Available at: https://www.ofcom.org.uk/__data/assets/pdf_file/0024/234609/childrens-media-use-and-attitudes-report-2022.pdf

[21] Cybersafe Kids, Academic Year in Review (2021-2022).

[22] The Common Sense Census: Media use by tweens and teens (common sense, 2021), Available at: https://www.commonsensemedia.org/sites/default/files/research/report/8-18-census-integrated-report-final-web_0.pdf

[23] NACOS: Report of a National Survey of Children, their Parents and Adults regarding Online Safety (2021)

[24] Ofcom, User experience of potential online harms within video sharing platforms. Available at: https://www.ofcom.org.uk/__data/assets/pdf_file/0024/216492/yonder-report-experience-of-potential-harms-vsps.pdf

[25] The U.S. Surgeon General's Advisory, Social Media and Youth Mental Health (2023). Available at: https://www.hhs.gov/sites/default/files/sg-youth-mental-health-social-media-advisory.pdf

[26] YONDER, Project 1 video-sharing platform usage & experience of harms survey (2020) Available at: https://www.ofcom.org.uk/__data/assets/pdf_file/0023/216518/yonder-report-experience-of-potential-harms-vsps.pdf

[27] Cybersafe Kids, Academic Year in Review (2021-2022).

[28] YONDER, Project 1 video-sharing platform usage & experience of harms survey (2020)

[29] Ofcom, User experience of potential online harms within video sharing platforms.

[30] YONDER, Project 1 video-sharing platform usage & experience of harms survey (2020)

[31] UNICEF, Cyberbullying: What is it and how to stop it (2023) Available at: https://www.unicef.org/end-violence/how-to-stop-cyberbullying

5.3 | Online content by which a person bullies or humiliates another person

[32] NSPCC, How children are being bullied (2016) Available at: https://library.nspcc.org.uk//HeritageScripts/Hapi.dll/search2?searchTerm0=C6189

[33] NSPCC, How children are being bullied (2016)

[34] ISPCC, A glossary of Cyberbullying Terms every parent should know about (2022). Available at: https://www.ispcc.ie/a-glossary-of-cyber-bullying-terms-every-parent-should-know-about/

[35] eSafety Commissioner, Young people > Trolling (2023). Available at: https://www.esafety.gov.au/young-people/trolling

[36] Cybersmile, Doxing (2023). Available at: https://www.cybersmile.org/advice-help/doxing

[37] Interrelate, Flaming (2023). Available at: https://www.interrelate.org.au/i-relate-teens/what-is-flaming#:~:text='Flaming'%20is%20posting%20personal%20insults,very%20aggressive%20form%20of%20intimidation.

[38] Internetmatters.org, Exploring the impacts of online harms Part 1 (2023) Available at: https://www.internetmatters.org/hub/news-blogs/inconsistency-between-parents-childrens-reports-online-harms/

[39] NACOS: Report of a National Survey of Children, their Parents and Adults regarding Online Safety (2021)

[40] Smahel et al, EU Kids Online 2020: Survey results from 19 countries (2020) Available at: https://www.lse.ac.uk/media-and-communications/assets/documents/research/eu-kids-online/reports/EU-Kids-Online-2020-10Feb2020.pdf

[41] Barnardos, Cyber Bullying Report (2022)

[42] YONDER, User experience of potential online harms within video sharing platforms (2021) Available at: https://www.ofcom.org.uk/__data/assets/pdf_file/0024/216492/yonder-report-experience-of-potential-harms-vsps.pdf

[43] Davidson et al, Research on Protection of Minors: A literature review and interconnected frameworks. Implications for VSP regulation and beyond (2021) Available at: https://www.ofcom.org.uk/__data/assets/pdf_file/0023/216491/uel-report-protection-of-minors.pdf

[44] Internetmatters.org, Insights from Internet Matters tracker survey (2022) Available at: https://www.internetmatters.org/wp-content/uploads/2023/02/Internet-Matters-Insights-Tracker-November-2022.pdf

[45] Office for National Statistics, Online bullying in England and Wales: year ending March 2020 (2020) Available at: https://www.ons.gov.uk/peoplepopulationandcommunity/crimeandjustice/bulletins/onlinebullyinginenglandandwales/yearendingmarch2020

[46] Barnardos, Cyber Bullying Report (2022)

[47] Better Internet for Kids, Classifying and responding to online risk to children (2023) Available at: https://www.betterinternetforkids.eu/documents/167024/200055/Good+practice+guide+-+Classifying+and+responding+to+online+risk+to+children+-+FINAL+-+February+2023.pdf

[48] Kostyrka-Allchorne et al, Review: Digital experiences and their impact on the lives of adolescents with pre-existing anxiety, depression, eating and non-suicidal self-injury conditions. (2022). Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10108198/

[49] Cambridge Dictionary, psychosomatic. Available at: https://dictionary.cambridge.org/dictionary/english/psychosomatic

[50] Davidson et al, Research on Protection of Minors: A literature review and interconnected frameworks. Implications for VSP regulation and beyond (2021)

[51] 5Rights Foundation, Risky-by-Design (2023) Available at: https://www.riskyby.design/introduction

[52] Patchin et al, Digital self-harm and suicidality among adolescents (2022) Available at: https://www.researchgate.net/publication/361114242_Digital_Self-Harm_and_Suicidality_Among_Adolescents

[53] 5Rights Foundation, Risky-by-Design (2023)

[54] 5Rights Foundation, Risky-by-Design (2023)

[55] Richardson, Milovidov, & Blamire (2017). Bullying: Perspectives, practice and insights. Council of Europe Available at: https://rm.coe.int/090000168078a78d

[56] 5Rights Foundation, Risky-by-Design (2023)

[57] NSPCC, Live streaming and video-chatting

[58] Ofcom, How Ofcom is building our evidence base around online fraud and illegal harms (2023) Available at: https://www.ofcom.org.uk/news-centre/2023/building-evidence-base-around-online-fraud-and-illegal-harms?utm_medium=email&utm_campaign=Weekly%20publications%20update%2026%20May%202023&utm_content=Weekly%20publications%20update%2026%20May%202023+CID_d3ae0f408f56b910afd52875f5e2c85b&utm_source=updates&utm_term=How%20Ofcom%20is%20building%20its%20evidence%20base%20around%20online%20fraud%20and%20illegal%20harms

[59] European Audiovisual Observatory, Mapping of media literacy practices and actions in EU-28 (2016) Available at: https://rm.coe.int/1680783500

[60] Finkelhor, Youth Internet Safety Education: Aligning programs with the evidence base (2020) Available at: https://journals.sagepub.com/doi/abs/10.1177/1524838020916257?journalCode=tvaa

[61] Milosevic, Changing the Paradigm for Cyberbullying Intervention and Prevention: Considering Dignity, Values, and Children's Rights (2021) Available at: https://www.ispcc.ie/guest-post-changing-the-paradigm-for-cyberbullying-intervention-and-prevention-considering-dignity-values-and-childrens-rights/

[62] Davidson et. al, Research on Protection of Minors: A literature review and interconnected frameworks. Implications for VSP regulation and beyond (2021)

[63] Department for Education, Cinealtas: Action Plan on Bullying (2022). Available at: https://www.gov.ie/pdf/?file=https://assets.gov.ie/241000/eb57d761-2963-4ab0-9d16-172b2e3be86d.pdf#page=null

[64] Anti-Bullying Centre, FUSE Impact Report 2022: FUSE Anti-bullying and online safety programme for primary and post-primary schools (2022) Available at: https://antibullyingcentre.ie/fuse/wp-content/uploads/2022/09/FUSE_22_ImpactReport.pdf

[65] For more information see: https://antibullyingcentre.ie/project/osaibi/

[66] Beat, The Dangers of Pro-Ana and Pro-Mia (2023), Available at: https://www.beateatingdisorders.org.uk/get-information-and-support/about-eating-disorders/dangers-of-pro-ana-and-pro-mia/

5.4 | Online content by which a person promotes or encourages behaviour that characterises a feeding or eating disorder

[67] Au, E., Cosh, S., Social media and eating disorder recovery: An exploration of Instagram recovery community users and their reasons for engagement (2022). Available at: https://www.sciencedirect.com/science/article/abs/pii/S1471015322000575?via%3Dihub

[68] Kostyrka-Allchorne et al, Review: Digital experiences and their impact on the lives of adolescents with pre-existing anxiety, depression, eating and non-suicidal self-injury conditions. (2022).

[69] Khatwa et al (2023) Young people and online eating disorder content: a qualitative evidence synthesis. London: EPPI Centre, UCL Social Research Institute, UCL Institute of Education, University College London.

[70] Padín, et al.,(2021). Social media and eating disorder psychopathology: A systematic review. Cyberpsychology: Journal of Psychosocial Research on Cyberspace, 15(3), Article 6. https://doi.org/10.5817/CP2021-3-6

[71] Aparicio-Martinez, et al. Social Media, Thin-Ideal, Body Dissatisfaction and Disordered Eating Attitudes: An Exploratory Analysis (2019). Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6861923/pdf/ijerph-16-04177.pdf

[72] Aparicio-Martinez, et al. Social Media, Thin-Ideal, Body Dissatisfaction and Disordered Eating Attitudes: An Exploratory Analysis (2019).

[73] Khatwa et al (2023) Young people and online eating disorder content: a qualitative evidence synthesis. London: EPPI Centre, UCL Social Research Institute, UCL Institute of Education, University College London.

[74] Au, E., Cosh, S., Social media and eating disorder recovery: An exploration of Instagram recovery community users and their reasons for engagement (2022).

[75] Khatwa et al (2023) Young people and online eating disorder content: a qualitative evidence synthesis. London: EPPI Centre, UCL Social Research Institute, UCL Institute of Education, University College London.

[76] Smahel et al, EU Kids Online 2020: Survey results from 19 countries (2020)

[77] Smahel et al, EU Kids Online 2020: Survey results from 19 countries (2020)

[78] NACOS: Report of a National Survey of Children, their Parents and Adults regarding Online Safety (2021)

[79] YONDER, User experience of potential online harms within video sharing platforms (2021)

[80] Centre for Countering Digital Hate, Deadly by Design. (2022) Available at: https://counterhate.com/wp-content/uploads/2022/12/CCDH-Deadly-by-Design_120922.pdf

[81] Farthing, Dr. R., Designing for Disorder: Inogram's pro-eating disorder bubble in Australia. (2022). Available at: https://au.reset.tech/uploads/insta-pro-eating-disorder-bubble-april-22-1.pdf

[82] YONDER, User experience of potential online harms within video sharing platforms (2021)

[83] Smahel et al, EU Kids Online 2020: Survey results from 19 countries (2020)

[84] Oksanen et al. (2016). Young people who access harm-advocating online content: A four-country survey. Cyberpsychology: Journal of Psychosocial Research on Cyberspace. Available at: https://cyberpsychology.eu/article/view/6179/5528

[85] Ioannidis et al. Problematic usage of the internet and eating disorder and related psychopathology: A multifaceted, systematic review and meta-analysis (2021)

[86] YONDER, User experience of potential online harms within video sharing platforms (2021)

[87] The U.S. Surgeon General's Advisory, Social Media and Youth Mental Health (2023).

[88] Zhang et al. The relationship between SNS usage and disordered eating behaviors: A meta-analysis. (2021) Frontiers in Psychology

[89] Academy for Eating Disorders, Urgent Responsibility to Reduce Harms Posed by Social Media on risk for Eating Disorders: An Open Letter to Facebook, Instagram, TikTok, and Other Global Social Media Corporations. (2021) Available at: https://www.newswise.com/articles/urgent-responsibility-to-reduce-harms-posed-by-social-media-on-risk-for-eating-disorders

[90] Cavazos-Rehg, P. et al. Examining the self-reported advantages and disadvantages of socially networking about body image and eating disorders (2020). Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8135099/

[91] Bodywhys, Statistics. Available at: https://www.bodywhys.ie/media-research/statistics/

[92] Bodywhys, Statistics.

[93] Saunders and Eaton, Snaps, Selfies and Shares: How three popular social media platforms contribute to the Sociocultural model of disordered eating among young women. Cyberpsychology, behaviour and social networking. (2018).

[94] Pruccoli et al., The use of TikTok among children and adolescents with Eating Disorders: experience in a third-level public Italian centre during the SARS-CoV-2 pandemic

[95] Bodywhys.ie, Lived experience – impact on people with eating disorders

[96] Padín, et al.,(2021). Social media and eating disorder psychopathology: A systematic review. Cyberpsychology: Journal of Psychosocial Research on Cyberspace, 15(3), Article 6. https://doi.org/10.5817/CP2021-3-6

[97] Cavazos-Rehg et al., Examining the self-reported advantages and disadvantages of socially networking about body image and eating disorders (2020)

[98] NSPCC (2022) Children's experiences of legal but harmful content online. Insight briefing. London: NSPCC.

[99] Perloff, R., Social media effects on young women's body image concerns: theoretical perspectives and an agenda for research (2014) Available at: https://is.muni.cz/el/1423/podzim2014/PSY221P121/um/Perloff2014.SocialMediaEffectsBodyImage.BID.pdf

[100] NSPCC (2022) Children's experiences of legal but harmful content online. Insight briefing. London: NSPCC.

[101] Perloff, R., Social media effects on young women's body image concerns: theoretical perspectives and an agenda for research (2014)

[102] Vandenbosch L,. Fardouly J., Tiggemann M,. Social media and body image: Recent trends and future directions (2022) Available at: https://www.sciencedirect.com/science/article/abs/pii/S2352250X21002414?via%3Dihub

[103] Karsay, K., Knoll, J., Matthes, J.,Sexualizing media use and self-objectification: a meta-analysis (2017) Available at: https://journals.sagepub.com/doi/full/10.1177/0361684317743019

[104] The Priory Group, The rise of digitally-altered photos on social media. Available at: https://www.priorygroup.com/blog/can-online-image-editing-on-social-media-contribute-to-eating-disorders

[105] Padín, et al.,(2021). Social media and eating disorder psychopathology: A systematic review. Cyberpsychology: Journal of Psychosocial Research on Cyberspace, 15(3), Article 6.

[106] Perloff, R., Social media effects on young women's body image concerns: theoretical perspectives and an agenda for research (2014)

[107] Khatwa et al (2023) Young people and online eating disorder content: a qualitative evidence synthesis. London: EPPI Centre, UCL Social Research Institute, UCL Institute of Education, University College London.

[108] Hockin-Boyers, et al 2020. Digital pruning: Agency and social media use as a personal political project among female weightlifters in recovery from eating disorders. New media & society, 23, 2345-2366.

[109] Khatwa et al (2023) Young people and online eating disorder content: a qualitative evidence synthesis. London: EPPI Centre, UCL Social Research Institute, UCL Institute of Education, University College London.

[110] Perloff, R., Social media effects on young women's body image concerns: theoretical perspectives and an agenda for research (2014)

[111] Khatwa et al (2023) Young people and online eating disorder content: a qualitative evidence synthesis. London: EPPI Centre, UCL Social Research Institute, UCL Institute of Education, University College London.

[112] Beat Eating Disorders, Eating Disorders and social media. Available at: https://www.beateatingdisorders.org.uk/your-stories/eating-disorders-and-social-media/

[113] Harriger et al., The dangers of the rabbit hole: Reflections on social media as a portal into a distorted world of edited bodies and eating disorder risk and the role of algorithms (2022). Available at: https://www.sciencedirect.com/science/article/abs/pii/S1740144522000638?via%3Dihub

[114] Padín, et al.,(2021). Social media and eating disorder psychopathology: A systematic review. Cyberpsychology: Journal of Psychosocial Research on Cyberspace, 15(3), Article 6.

[115] Harriger et al., The dangers of the rabbit hole: Reflections on social media as a portal into a distorted world of edited bodies and eating disorder risk and the role of algorithms (2022)

[116] 5Rights Foundation, Risky-by-Design (2023)

[117] 5Rights Foundation, Risky-by-Design (2023)

[118] 5Rights Foundation, Risky-by-Design (2023)

[119] NSPCC, Live streaming and video-chatting

[120] Samaritans. Implementing effective moderation for self-harm and suicide content. Available at: https://media.samaritans.org/documents/Implementing_effective_moderation_for_self-harm_and_suicide_content_FINAL.pdf

[121] Gerrard, Y. How we're helping social media companies remove harmful content and protect their users. Available at: https://www.sheffield.ac.uk/research/features/social-media-and-mental-health

[122] Academy for Eating Disorders, Urgent Responsibility to Reduce Harms Posed by Social Media on risk for Eating Disorders: An Open Letter to Facebook, Instagram, TikTok, and Other Global Social Media Corporations. (2021)

[123] Harriger et al., The dangers of the rabbit hole: Reflections on social media as a portal into a distorted world of edited bodies and eating disorder risk and the role of algorithms (2022)

[124] Harriger et al., The dangers of the rabbit hole: Reflections on social media as a portal into a distorted world of edited bodies and eating disorder risk and the role of algorithms (2022)

[125] Bodywhys. Available at: https://www.bodywhys.ie

[126] Mcternan, N., Ryan, F. The Harmful impact of suicide and self-harm content online: A review of the literature (2020) Available at: https://www.nsrf.ie/wp-content/uploads/2021/04/Harmful-impact-of-suicide-and-self-harm-content-online-Review-of-the-literature-Final.pdf

5.5 | Online content by which a person promotes or encourages self-harm or suicide, or makes available knowledge of relevant methods

[127] NACOS: Report of a National Survey of Children, their Parents and Adults regarding Online Safety (2021)

[128] Samaritans, How social media users experience self-harm and suicide content (2022). Available at: https://media.samaritans.org/documents/Samaritans_How_social_media_users_experience_self-harm_and_suicide_content_WEB_v3.pdf

[129] Arendt, F. Suicide rate and information seeking via search engines: A cross-national correlational approach. (2018). Available at: https://pubmed.ncbi.nlm.nih.gov/29173051/

[130] Lup et al, Instagram #instasad?: Exploring associations among instagram use, depressive symptoms, negative social comparison, and strangers followed. (2015). Cyberpsychology, Behavior and Social Networking. Available at: https://pubmed.ncbi.nlm.nih.gov/25965859/

[131] Gould et al., Psychopathology associated with suicidal ideation and attempts among children and adolescents. (1998). Journal of the American Academy of Child and Adolescent Psychiatry. Available at: https://pubmed.ncbi.nlm.nih.gov/9735611/

[132] Reinherz et al., Early psychosocial risks for adolescent suicidal ideation and attempts. (1995). Journal of the American Academy of Child and Adolescent Psychiatry. Available at: https://pubmed.ncbi.nlm.nih.gov/7775355/

[133] HSE National Office for Suicide Prevention, Annual Report 2021. Available at: https://www.hse.ie/eng/services/list/4/mental-health-services/nosp/about/annualreports/

[134] World Health Organization. (2019). Suicide. Available from: https://www.who.int/news-room/fact-sheets/detail/suicide

[135] Hawton et al., Mortality in children and adolescents following presentation to hospital after non-fatal self-harm in the multicentre study of self-harm: A prospective observational cohort study. (2020). The Lancet Child and Adolescent Health. Available at: https://pubmed.ncbi.nlm.nih.gov/31926769/

[136] Financial Times, The teen mental health crisis: a reckoning for Big Tech (2023). Available at: https://www.ft.com/content/77d06d3e-2b9f-4d46-814f-da2646fea60c?accessToken=zwAF_niZUlZQkc930G0-K59NRtOBT9omRv6mDA.MEUCIQCr834OcktSR5nuA7hG9alMFK__YNwOjD3FEMNFXiGetgIgJUBUare8wMwhkEG8PiFGLMpvUkO2a2S76qrGpCF6en0&sharetype=gift&token=ded52684-d859-49f8-9f58-b4198db423e8

[137] Samaritans, How social media users experience self-harm and suicide: 2022

[138] Alphr, Facebook And YouTube Are The Worst Offenders For Exposing Children To Suicide And Sex, NSPCC Report Finds Available at: https://www.alphr.com/social-media/1009387/nspcc-youtube-facebook-adult-content/

[139] NSPCC, Content promoting self-harm, suicide and eating disorders (2021) https://www.nspcc.org.uk/keeping-children-safe/online-safety/inappropriate-explicit-content/promotion-self-harm/

[140] Marchant, A. et al., Impact of Web-Based Sharing and Viewing of Self-Harm-Related Videos and Photographs on Young People: Systematic Review. (2021) https://pubmed.ncbi.nlm.nih.gov/33739289/

[141] Samaritans, How social media users experience self-harm and suicide: 2022

[142] The U.S. Surgeon General's Advisory, Social Media and Youth Mental Health (2023).

[143] NSPCC, Content promoting self-harm, suicide and eating disorders (2021) https://www.nspcc.org.uk/keeping-children-safe/online-safety/inappropriate-explicit-content/promotion-self-harm/

[144] Coroner Report (2022), available at: https://www.judiciary.uk/wp-content/uploads/2022/10/Molly-Russell-Prevention-of-future-deaths-report-2022-0315_Published.pdf

[145] Davidson et al., Research on Protection of Minors, (2021).

[146] Singaravelu et al., Information-Seeking on the Internet . (2015) Available at: https://pubmed.ncbi.nlm.nih.gov/26088826/

[147] Samaritans, How social media users experience self-harm and suicide (2022)

[148] Davidson et al., Research on Protection of Minors, (2021).

[149] Griffith, Erin; Metz, Cade (2023-01-27). "Anthropic Said to Be Closing In on $300 Million in New A.I. Funding". The New York Times. Retrieved 2023-03-14.

[150] Brown, R. C. et al. #cutting: Non-suicidal self-injury (NSSI) on Instagram (2018) Available at: https://pubmed.ncbi.nlm.nih.gov/28705261/

[151] Centre for Countering Digital Hate, Deadly by Design. (2022)

[152] Daily Mail, Teenager, 17, who simply 'liked' some sad quotes on Instagram reveals how the site's algorithm sucked her into suicide groups - and admits it made her believe self-harm was 'glamorous' (2022). Available at: https://www.dailymail.co.uk/femail/article-10485227/Facebook-whistleblower-Frances-Haugen-warns-teens-killing-Instagram.html

[153] Gerrard, Y. Beyond the hashtag: Circumventing content moderation on social media (2018) Available at: https://journals.sagepub.com/doi/abs/10.1177/1461444818776611?journalCode=nmsa

[154] Lup et al, Instagram #instasad?: Exploring associations among instagram use, depressive symptoms, negative social comparison, and strangers followed. (2015). Cyberpsychology, Behavior and Social Networking.

[155] Future Care Capital, Three-quarters of social media users see self-harm content online by age 14 (2022). Available at: https://futurecarecapital.org.uk/latest/three-quarters-users-see-self-harm-content/

[156] Murphy, The Top Mental Health TikTok Influencers and Why They're Important (2021) Available at: https://www.everydayhealth.com/emotional-health/under-pressure/the-top-mental-health-tiktok-influencers-and-why-theyre-important/

[157] Definition of gratuitous violence (Collins Dictionary, 2023). Available at: https://www.collinsdictionary.com/dictionary/english/gratuitous-violence

5.6 | Other online content which impairs the physical, mental, or moral development of minors

[158] Definition of sexually explicit (LawInsider, 2023). Available at: https://www.lawinsider.com/dictionary/sexually-explicit

[159] Au, E., Cosh, S., Social media and eating disorder recovery: An exploration of Instagram recovery community users and their reasons for engagement (2022).

[160] Khatwa et al (2023) Young people and online eating disorder content: a qualitative evidence synthesis. London: EPPI Centre, UCL Social Research Institute, UCL Institute of Education, University College London.

[161] Smahel et al, EU Kids Online 2020: Survey results from 19 countries (2020)

[162] Report of a National Survey of Children, their Parents and Adults regarding Online Safety (National Advisory Council for Online Safety, 2021). Available at: https://www.gov.ie/pdf/?file=https://assets.gov.ie/204409/b9ab5dbd-8fdc-4f97-abfc-a88afb2f6e6f.pdf#page=null

[163] Report of a National Survey of Children, their Parents and Adults regarding Online Safety (National Advisory Council for Online Safety, 2021).

[164] EU Kids Online 2020: Survey results from 19 countries (Smahel et al, 2020).

[165] ‘A lot of it is actually just abuse’: Young people and pornography (Children’s Commissioner, 2023). Available at: https://assets.childrenscommissioner.gov.uk/wpuploads/2023/02/cc-a-lot-of-it-is-actually-just-abuse-young-people-and-pornography-updated.pdf

[166] EU Kids Online 2020: Survey results from 19 countries (Smahel et al, 2020).

[167] Not including terrorist content. This topic is covered in Section 3.8.

[168] Report of a National Survey of Children, their Parents and Adults regarding Online Safety (National Advisory Council for Online Safety, 2021). Available at: https://www.gov.ie/pdf/?file=https://assets.gov.ie/204409/b9ab5dbd-8fdc-4f97-abfc-a88afb2f6e6f.pdf#page=null

[169] Report of a National Survey of Children, their Parents and Adults regarding Online Safety (National Advisory Council for Online Safety, 2021).

[170] Pornography regulation: the case for Parliamentary reform (APPG on Commercial Sexual Exploitation, 2023). Available at: https://www.appg-cse.uk/wp-content/uploads/2023/02/Inquiry-on-pornography.pdf

[171] Social Media Animal Cruelty Coalition, Making Money from Misery (2021). Available at: https://www.smaccoalition.com/smacc-report

[172] Social Media Animal Cruelty Coalition, Making Money from Misery (2021).

[173] Reports of animal abuse on social media has more than doubled in the last year (RSPCA, 2022). Available at: https://www.rspca.org.uk/-/news-reports-of-animal-abuse-doubled

[174] 23% of kids have seen animal abuse online (RSPCA, 2018). Available at: https://www.rspca.org.uk/-/16_10_18_genkind

[175] Reports of animal abuse on social media has more than doubled in the last year (RSPCA, 2022). Available at: https://www.rspca.org.uk/-/news-reports-of-animal-abuse-doubled

[176] 23% of kids have seen animal abuse online (RSPCA, 2018). Available at: https://www.rspca.org.uk/-/16_10_18_genkind

[177] The contents of the table were mainly taken from the 'A lot of it is actually just abuse': Young people and pornography (2023) report by the UK Children's Commissioner, unless stated otherwise.

[178] 23% of kids have seen animal abuse online (RSPCA, 2018).

[179] EU Kids Online 2020: Survey results from 19 countries (2020). Available at: https://www.lse.ac.uk/media-and-communications/assets/documents/research/eu-kids-online/reports/EU-Kids-Online-2020-10Feb2020.pdf

[180] NSPCC, Children's experiences of legal but harmful content online (2022). Available at: https://learning.nspcc.org.uk/media/2727/legal-but-harmful-content-online-helplines-insight-briefing.pdf

[181] The role of callous/unemotional traits in mediating the association between animal abuse exposure and behavior problems among children exposed to intimate partner violence (McDonald et al., 2017). Available at: https://www.sciencedirect.com/science/article/abs/pii/S0145213417303356?via%3Dihub

[182] NSPCC, Children's experiences of legal but harmful content online (2022).

[183] EU Kids Online 2020: Survey results from 19 countries (Smahel et al, 2020).

[184] Child 'self-generated' sexual material online: Children and young people's perspectives' (WeProtect Global Alliance, 2023). Available at: https://praesidiosafeguarding.co.uk/safe-guarding/uploads/2023/05/WPReportPraesidioSafeguarding.pdf?x70166

[185] A lot of it is actually just abuse': Young people and pornography (Children's Commissioner, 2023).

[186] EU Kids Online 2020: Survey results from 19 countries (Smahel et al, 2020).

[187] NSPCC, Children's experiences of legal but harmful content online (2022).

[188] Pornography regulation: the case for Parliamentary reform (APPG on Commercial Sexual Exploitation, 2023). Available at: https://www.appg-cse.uk/wp-content/uploads/2023/02/Inquiry-on-pornography.pdf

[189] 'A lot of it is actually just abuse': Young people and pornography (Children's Commissioner, 2023).

[190] NSPCC, Children's experiences of legal but harmful content online (2022).

[191] 'A lot of it is actually just abuse': Young people and pornography (Children's Commissioner, 2023). Available at: https://assets.childrenscommissioner.gov.uk/wpuploads/2023/02/cc-a-lot-of-it-is-actually-just-abuse-young-people-and-pornography-updated.pdf

[192] 'A lot of it is actually just abuse': Young people and pornography (Children's Commissioner, 2023).

[193] European Parliament, The impact of the use of social media on women and girls (2023). Available at: https://www.europarl.europa.eu/RegData/etudes/STUD/2023/743341/IPOL_STU(2023)743341_EN.pdf

[194] Independent.ie, Exposure to pornography leading to 'increased levels of sexualised behaviour among children' (2023). Available at: https://www.independent.ie/regionals/wexford/wexford-district/exposure-to-pornography-leading-to-increased-levels-of-sexualised-behaviour-among-children/a1740793130.html

[195] The Lancet, Media violence and youth aggression (2017). Available at: https://www.thelancet.com/journals/lanchi/article/PIIS2352-4642(17)30033-0/fulltext

[196] The Lancet, Media violence and youth aggression (2017).

[197] The Lancet, Media violence and youth aggression (2017).

[198] BBC, 'Do video games make people violent?' (2015). Available at: https://journals.sagepub.com/doi/pdf/10.1177/1461444821994490https://www.bbc.co.uk/news/technology-33960075

[199] Guardian, Playing video games doesn't lead to violent behaviour, study shows (2020). Available at: https://www.theguardian.com/games/2020/jul/22/playing-video-games-doesnt-lead-to-violent-behaviour-study-shows

[200] Allardyce and Yates, Working with young people who have displayed harmful sexual behaviour (2020).

[201] New Media & Society, Gangs and social media: a systematic literature review and an identification of future challenges, risks, and recommendations (2021). Available at:

[202] The Sun, Vicious street gang war played out on Instagram and Youtube jail videos' (2018). Available at: https://www.thesun.co.uk/news/6686578/vicious-street-gang-war-played-out-on-instagram-and-youtube-jail-videos/

[203] The Future of Children, The impact of violence on children (1999). Available at: https://www.semanticscholar.org/paper/The-impact-of-violence-on-children.-Osofsky/fe42c5ecd574bfdcab03b350e9664fabc054c83a.

[204] 5rights Foundation, Pathways: How digital design puts children at risk (2021). Available at: https://5rightsfoundation.com/uploads/Pathways-how-digital-design-puts-children-at-risk.pdf

[205] Ionos, Upload filters: a danger to free internet content? (2019). Available at: https://www.ionos.com/digitalguide/websites/digital-law/upload-filters/

[206] Ionos, Upload filters: a danger to free internet content? (2019).

[207] A lot of it is actually just abuse': Young people and pornography (Children's Commissioner, 2023).

[208] Family Online Safety Institute, Three Reasons Social Media Age Restrictions Matter, (2023). Available at: https://www.fosi.org/good-digital-parenting/three-reasons-social-media-age-restrictions-matter#

[209] Davidson et al, Research on Protection of Minors: A literature review and interconnected frameworks. Implications for VSP regulation and beyond (2021)

[210] Davidson et al, Research on Protection of Minors: A literature review and interconnected frameworks. Implications for VSP regulation and beyond (2021)

[211] Young people, Pornography & Age-verification (BBFC, 2020). Available at: https://revealingreality.co.uk/wp-content/uploads/2020/01/BBFC-Young-people-and-pornography-Final-report-2401.pdf

[212] A lot of it is actually just abuse': Young people and pornography (Children's Commissioner, 2023).

[213] Young people, Pornography & Age-verification (BBFC, 2020).

[214] How Do Social Media Algorithms Work? (Digital Marketing Institute, 2023). Available at: https://digitalmarketinginstitute.com/blog/how-do-social-media-algorithms-work

[215] Twitter, About age screening on Twitter (2023). Available at: https://help.twitter.com/en/safety-and-security/age-verification#

[216] Young people, Pornography & Age-verification (BBFC, 2020).

[217] Young people, Pornography & Age-verification (BBFC, 2020).

[218] Young people, Pornography & Age-verification (BBFC, 2020).

[219] The 5Rights Foundation, But how do they know it is a child? (2021). Available at: https://5rightsfoundation.com/uploads/But_How_Do_They_Know_It_is_a_Child.pdf

[220] Arstechnica, Seven states push to require ID for watching porn online (2023). Available at: https://arstechnica.com/tech-policy/2023/02/seven-states-push-to-require-id-for-watching-porn-online/

[221] Dame Rachel de Souza: Children's Commissioner Annual Report 2021-22: Available at https://www.childrenscommissioner.gov.uk/corporate-governance/annual-report-2021-22/

[222] Bryant Furlow: Media violence and youth aggression, The Lancet (2017) Available at : https://www.thelancet.com/journals/lanchi/article/PIIS2352-4642(17)30033-0/fulltext

[223] The Lancet, Media violence and youth aggression (2017).

[224] The Lancet, Media violence and youth aggression (2017).

[225] European Union Agency for Fundamental Rights, Article 21 – Non-Discrimination (). Available at: https://fra.europa.eu/en/eu-charter/article/21-non-discrimination#

5.7 | Online content by which a person incites hatred or violence

[226] Amnesty International, Toxic Twitter – Women's Experiences Of Violence and Abuse On Twitter (2023). Available at: https://www.amnesty.org/en/latest/news/2018/03/online-violence-against-women-chapter-3-2/

[227] Irish Council for Civil Liberties, Lifecycle of a Hate Crime (2018). Available at: https://www.iccl.ie/wp-content/uploads/2018/04/Life-Cycle-of-a-Hate-Crime-Country-Report-for-Ireland.pdf

[228] Council on Foreign Relations, Hate Speech on Social Media: Global Comparisons (2019). Available at: https://www.cfr.org/backgrounder/hate-speech-social-media-global-comparisons

[229] Ribeiro et al, The evolution of the manosphere across the web (2021). Available at: https://ojs.aaai.org/index.php/ICWSM/article/view/18053

[230] The Guardian, Inside the Violent, Misogynistic world of Tiktok's new star, Andrew Tate (2022). Available at: https://www.theguardian.com/technology/2022/aug/06/andrew-tate-violent-misogynistic-world-of-tiktok-new-star

[231] Business Insider India, TikTok 'influencers' charged for hate speech and attempting to incite communal violence (2019). Available at: https://www.businessinsider.in/tiktok-influencers-charged-for-hate-speech-and-attempting-to-incite-communal-violence/articleshow/70157769.cms

[232] Sugiura, The Incel Rebellion: The Rise of the Manosphere and the Virtual War against Women (2021). Available at: https://www.emerald.com/insight/content/doi/10.1108/978-1-83982-254-420211004/full/html#ch02-lev1-4

[233] Ofcom, User experience of Potential Online Harms within Video Sharing Platforms (2021).

[234] Ofcom (2018) Adults' Media Use and Attitudes. London: Ofcom. Available at: https://www.ofcom.org.uk/__data/assets/pdf_file/0011/113222/Adults-Media-Use-andAttitudes-Report-2018.pdf

[235] Ofcom (2018) Adults' Media Use and Attitudes. London: Ofcom.

[236] Ofcom (2018) Adults' Media Use and Attitudes. London: Ofcom.

[237] The U.S. Surgeon General's Advisory, Social Media and Youth Mental Health (2023).

[238] Ditch the Label, Uncovered: Online Hate Speech in the Covid Era. Available at: https://dtl-beta-website-assets.s3.amazonaws.com/Uncovered_Online_Hate_Speech_DT_Lx_BW_V2_1_6aef9e5c5f.pdf

[239] SELMA, Hacking Online Hate: Building an evidence base for Educators (2019). Available at: https://hackinghate.eu/assets/documents/hacking-online-hate-research-report-1.pdf

[240] SELMA, Hacking Online Hate: Building an evidence base for Educators (2019).

[241] iReport.ie, Reports of racism in Ireland (2022). Available at: https://inar.ie/wp-content/uploads/2021/03/2020_iReport.pdf

[242] iReport.ie, Reports of racism in Ireland (2022).

[243] iReport.ie, Reports of racism in Ireland (2022).

[244] Matthew L Williams et al: Hate in the Machine: Anti-Black and Anti-Muslim Social Media Posts as Predictors of Offline Racially and Religiously Aggravated Crime. (2019) Available at https://academic.oup.com/bjc/article/60/1/93/5537169

[245] Littler, M. & Feldman, M., Tell MAMA Reporting 2014/2015: Annual Monitoring, Cumulative Extremism, and Policy Implications. Teeside: Teesside University Press. (2015) Available at: https://www.tellmamauk.org/wpcontent/uploads/pdf/Tell%20MAMA%20Reporting%202014-2015.pdf

[246] Littler, M. & Feldman, M., Tell MAMA Reporting 2014/2015: Annual Monitoring, Cumulative Extremism, and Policy Implications. Teeside: Teesside University Press. (2015)

[247] UN Women & World Health Organisation, Technologically-facilitated violence against women: taking stock of evidence and data collection (2023). Available at: https://www.unwomen.org/sites/default/files/2023-04/Technology-facilitated-violence-against-women-Taking-stock-of-evidence-and-data-collection-en.pdf

[248] Women's Aid, One in five (2020).

[249] Women's Aid, One in five (2020).

[250] The Economist, Measuring the prevalence of online violence against women (2021). Available at: https://onlineviolencewomen.eiu.com/

[251] Amnesty International, Unsocial media; the real toll of online abuse against women (2017). Available at https://medium.com/amnesty-insights/unsocial-media-the-real-toll-of-online-abuse-against-women-37134ddab3f4

[252] Childnet, Project deShame Report (2017). Available at: https://www.childnet.com/what-we-do/our-projects/project-deshame/research/

[253] UN Women, Online and ICT facilitated violence against women and girls during COVID-19 (2020). Available at: https://www.unwomen.org/sites/default/files/Headquarters/Attachments/Sections/Library/Publications/2020/Brief-Online-and-ICT-facilitated-violence-against-women-and-girls-during-COVID-19-en.pdf

[254] Centre on Gender Equity and Health University of California San Diego, Trends in online misogyny before and during the COVID-19 pandemic: Analysis of Twitter data from five South-Asian countries (2021). Available at: https://data2x.org/wp-content/uploads/2021/03/UCSD-Brief-3_BigDataGenderCOVID19SouthAsianMisogyny.pdf

[255] The Economist, Measuring the prevalence of online violence against women (2021).

[256] United Nations Education, Scientific and Cultural Organization, Online violence Against Women Journalists: A Global Snapshot of Incidence and Impacts (2020). Available at: https://www.icfj.org/sites/default/files/2020-12/UNESCO%20Online%20Violence%20Against%20Women%20Journalists%20-%20A%20Global%20Snapshot%20Dec9pm.pdf

[257] All-Party Parliamentary Group, Pornographic regulation, The Case for Parliamentary reform (2023). Available at: https://www.appg-cse.uk/wp-content/uploads/2023/02/Inquiry-on-pornography.pdf

[258] UN Women et al., Accelerating efforts to tackle online and technology facilitated violence against women and girls (2022). Available at: https://www.unwomen.org/sites/default/files/2022-10/Accelerating-efforts-to-tackle-online-and-technology-facilitated-violence-against-women-and-girls-en_0.pdf

[259] United Nations Education, Scientific and Cultural Organization, Online violence Against Women Journalists: A Global Snapshot of Incidence and Impacts (2020). Available at: https://www.icfj.org/sites/default/files/2020-12/UNESCO%20Online%20Violence%20Against%20Women%20Journalists%20-%20A%20Global%20Snapshot%20Dec9pm.pdf

[260] The Economist, Measuring the prevalence of online violence against women (2021). Available at: https://onlineviolencewomen.eiu.com

[261] Dublin City University, Social media and online hostility: Experiences of women in Irish journalism (2023). Available at: https://www.dcu.ie/commsteam/policy/social-media-and-online-hostility-experiences-women-irish-journalism

[262] UN Women et al., Accelerating efforts to tackle online and technology facilitated violence against women and girls (2022). Available at: https://www.unwomen.org/sites/default/files/2022-10/Accelerating-efforts-to-tackle-online-and-technology-facilitated-violence-against-women-and-girls-en_0.pdf

[263] UNESCO, The Chilling: Global trends in online violence against women journalists (2021). Available at: https://en.unesco.org/sites/default/files/the-chilling.pdf

[264] The Economist, Measuring the prevalence of online violence against women (2021). Available at: https://onlineviolencewomen.eiu.com/

[265] The Economist, Measuring the prevalence of online violence against women (2021).

[266] Stonewall, The School Report (2017). Available at: https://www.stonewall.org.uk/system/files/the_school_report_2017.pdf

[267] Hubbard, A Survey of Online Anti-LGBT+ Hate Speech and Hate Crime (2020). Available at: https://safetobe.eu/wp-content/uploads/2020/10/Survey-online-hate-crimes-report.pdf

[268] Stonewall, The School Report (2017).

[269] European Commission, The Rise of Antisemitism Online during the Pandemic (2021). Available at: https://op.europa.eu/en/publication-detail/-/publication/d73c833f-c34c-11eb-a925-01aa75ed71a1/language-en

[270] European Union Agency for Fundamental Rights, Experiences and perceptions of antisemitism – second survey on discrimination and hate crime against Jew in the EU – summary (2019). Available at: https://fra.europa.eu/en/publication/2018/experiences-and-perceptions-antisemitism-second-survey-discrimination-and-hate

[271] Centre for Countering Digital Hate, Failure to protect: How tech giants fail to act on user reports of antisemitism. (2021) Available at: https://counterhate.com/wp-content/uploads/2022/05/Failure-to-Protect.pdf

[272] Centre for Countering Digital Hate, Failure to protect: How tech giants fail to act on user reports of antisemitism. (2021)

[273] Anti-Muslim Hate - Failure To Protect Report (2022). Available at: http://counterhate.com/research/anti-muslim-hate/

[274] The Institute for Strategic Dialogue, Hatescape: An In-Depth Analysis of Extremism and Hate Speech on TikTok (2021). Available at: https://www.isdglobal.org/wp-content/uploads/2021/08/HateScape_v5.pdf

[275] Global Witness: Facebook approves ads inciting violence in Northern Ireland (2021) Available at : https://www.euractiv.com/section/media/news/ngo-facebook-approved-ads-inciting-violence-in-n-ireland/

[276] Office for Democratic Institutions and Human Rights, ODIHR's Hate Crime Data for 2021 (2021). Available at: https://hatecrime.osce.org/sites/default/files/2022-11/2021%20Hate%20Crime%20Data%20Findings%20-%20presentation_161122.pdf

[277] Cardiff University, Increase in online hate speech leads to more crimes against minorities (2019). Available at: https://www.cardiff.ac.uk/news/view/1702622-increase-in-online-hate-speech-leads-to-more-crimes-against-minorities

[278] Cardiff University, Increase in online hate speech leads to more crimes against minorities (2019).

[279] Cardiff University, Increase in online hate speech leads to more crimes against minorities (2019).

[280] Social Development Direct, Technology-facilitated gender based violence (2022). Available at: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1168818/Technology_facilitated_gender_based_violence_preliminary_landscape_analysis.pdf

[281] Business Insider, TikTok 'influencers' charged for hate speech and attempting to incite communal violence (2019). Available at: https://www.businessinsider.in/tiktok-influencers-charged-for-hate-speech-and-attempting-to-incite-communal-violence/articleshow/70157769.cms

[282] The Economist, Measuring the prevalence of online violence against women (2020). Available at: https://onlineviolencewomen.eiu.com/

[283] Plan International, Free To Be Online? Girls' and young womens' experiences of online harassment (2020). Available at: https://plan-international.org/uploads/2022/02/sotwgr2020-execsummary-en-3.pdf

[284] Australian e-Safety Commissioner, Women in the Spotlight – Women's experiences with online abuse in their working lives (2022). Available at: https://www.esafety.gov.au/sites/default/files/2022-02/WITS-Womens-experiences-with-online-abuse-in-their-working-lives_0.pdf

[285] Plan International, Free To Be Online? Girls' and young women's' experiences of online harassment (2020). Available at: https://plan-international.org/uploads/2022/02/sotwgr2020-execsummary-en-3.pdf

[286] The Economist, Measuring the prevalence of online violence against women (2021).

[287] The Economist, Measuring the prevalence of online violence against women (2020). Available at: https://onlineviolencewomen.eiu.com/

[288] NSPCC, Radicalisation (2021). Available at: https://learning.nspcc.org.uk/safeguarding-child-protection/radicalisation#skip-to-content

[289] Stonewall, The School Report (2017).

[290] Stonewall, The School Report (2017).

[291] Stonewall, The School Report (2017).

[292] iReport:Reports of Racism in Ireland: (2020), Available at: https://inar.ie/wp-content/uploads/2021/03/2020_iReport-Reports-of-Racism-in-Ireland.pdf

[293] Institute of for Strategic Dialogue, Hatescape: An In-Depth Analysis of Extremism and Hate Speech on TikTok. (2021). Available at: https://www.isdglobal.org/wp-content/uploads/2021/08/HateScape_v5.pdf

[294] CBS news, Facebook whistleblower says company incentivizes "angry, polarizing, divisive content (2021). Available at: https://www.cbsnews.com/news/facebook-whistleblower-frances-haugen-60-minutes-polarizing-divisive-content/

[295] Mark Zuckerberg, A Blueprint for Content Governance and Enforcement. Available at: https://m.facebook.com/nt/screen/?params=%7B%22note_id%22%3A751449002072082%7D&path=%2Fnotes%2Fnote%2F&_rdr=&ref=buffer.com

[296] Facebook ad targeting could be used to incite sectarian violence, Global Witness Organisation (2021). Available at: https://www.globalwitness.org/en/press-releases/facebook-ad-targeting-could-be-used-incite-sectarian-violence/

[297] The Irish Times, Hateful ads submitted to Facebook, TikTok and YouTube approved by platforms (2023). Available at: https://www.irishtimes.com/technology/big-tech/2023/02/23/hateful-ads-submitted-to-facebook-tiktok-and-youtube-approved-by-platforms/

[298] Galop, Online Hate Crime Report (2020). Available at: https://www.report-it.org.uk/files/online-crime-2020_0.pdf

[299] Galop, Online Hate Crime Report (2020).

[300] Galop, Online Hate Crime Report (2020).

[301] Ofcom (2018) Adults' Media Use and Attitudes. London: Ofcom.

[302] Ofcom (2018) Adults' Media Use and Attitudes. London: Ofcom.

[303] SELMA, Hacking Online Hate: Building an evidence base for Educators (2019).

[304] SELMA, Hacking Online Hate: Building an evidence base for Educators (2019).

[305] The Economist, Measuring the prevalence of online violence against women (2021).

[306] The Economist, Measuring the prevalence of online violence against women (2021).

[307] Osservatorio Balcani Caucaso Transeuropa, Hate speech: what it is and how to contrast it (2018). Available at: https://www.rcmediafreedom.eu/Dossiers/Hate-speech-what-it-is-and-how-to-contrast-it

[308] UN Women, Online and ICT* facilitated violence against women and girls during COVID-19 (2020). Available at: https://www.unwomen.org/sites/default/files/Headquarters/Attachments/Sections/Library/Publications/2020/Brief-Online-and-ICT-facilitated-violence-against-women-and-girls-during-COVID-19-en.pdf

[309] UN Women & World Health Organisation, Technologically-facilitated violence against women: taking stock of evidence and data collection (2023).

[310] Rape victim sues Idaho lawmakers for outing and harassing her (Pettersson, 2023). Available at: https://www.courthousenews.com/rape-victim-sues-idaho-lawmakers-for-outing-and-harassing-her/

5.8 | Offences relating to the online identification of victims, suspects, or vulnerable people

[311] The University of Idaho Murders Show the Hidden Cost of America's True Crime Addiction (Time, 2023). Available at: https://time.com/6250031/idaho-murders-true-crime-addiction/

[312] Mason Greenwood: Manchester United footballer further arrested on suspicion of sexual assault and threats to kill (SkyNews, 2022). Available at: https://news.sky.com/story/mason-greenwood-manchester-united-footballer-further-arrested-on-suspicion-of-sexual-assault-and-threats-to-kill-12530505

[313] These impacts were taken from the case studies mentioned in the 'Prevalence and risk of harm' section prior.

[314] The University of Idaho Murders Show the Hidden Cost of America's True Crime Addiction (Time, 2023). Available at: https://time.com/6250031/idaho-murders-true-crime-addiction/

[315] Rape victim sues Idaho lawmakers for outing and harassing her (Pettersson, 2023). Available at: https://www.courthousenews.com/rape-victim-sues-idaho-lawmakers-for-outing-and-harassing-her/

[316] 'My own form of justice': rape survivors and the risk of social media 'vigilantism' (The Guardian, 2016). Available at: https://www.theguardian.com/society/2016/sep/13/social-media-rape-survivors-justice-legal-system

[317] The University of Idaho Murders Show the Hidden Cost of America's True Crime Addiction (Time, 2023).

[318] Idaho student murders: Rumours, 'clues' and online detectives (BBC, 2022). Available at: https://www.bbc.co.uk/news/world-us-canada-63840486

5.9 | Online content associated with terrorism

[319] Terrorism Situation and Trend Report (Europol, 2022). Available at: https://www.europol.europa.eu/cms/sites/default/files/documents/Tesat_Report_2022_0.pdf

[320] How did a former Irish soldier end up a member of Isis? (The Irish Times, 2022). Available at: https://www.irishtimes.com/crime-law/courts/2022/05/30/full-story-of-the-lisa-smith-trial-how-did-a-member-of-irish-defence-forces-end-up-a-member-of-isis/

[321] Ofcom, Ofcom: The Buffalo attack: Implications for online safety (2022). Available at: https://www.counterterrorism.police.uk/together-were-tackling-online-terrorism/

[322] Vox, ISIS videos are sickening. They're also really effective (2015). Available at: https://www.vox.com/videos/2015/7/6/8886461/isis-videos-burning

[323] The Guardian, Facebook trained its AI to block violent live streams after Christchurch attacks (2021). Available at: https://www.theguardian.com/technology/2021/oct/29/facebook-trained-its-ai-to-block-violent-live-streams-after-christchurch-attacks

[324] The Guardian, Facebook trained its AI to block violent live streams after Christchurch attacks (2021). Available at: https://www.theguardian.com/technology/2021/oct/29/facebook-trained-its-ai-to-block-violent-live-streams-after-christchurch-attacks

[325] European Union, Terrorism Situation and Trend Report, (2021). Available at: https://www.europol.europa.eu/cms/sites/default/files/documents/tesat_2021_0.pdf

[326] Dr. Kenyon, J., Exploring the role of the Internet in radicalisation and offending of convicted extremists (2021). Available at: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1017413/exploring-role-internet-radicalisation.pdf

[327] Euractiv: Regulating against radicalisation: Up to 400 online platforms hosting terrorist content (2019). Available at: https://www.euractiv.com/section/cybersecurity/news/up-to-400-online-platforms-hosting-terrorist-content-commission-says/

[328] European Parliament Briefing: Addressing the dissemination of terrorist content online Proposal for a regulation of the European Parliament and of the Council on preventing the dissemination of terrorist content online, (2018). Available at: https://www.europarl.europa.eu/RegData/etudes/BRIE/2020/649326/EPRS_BRI(2020)649326_EN.pdf

[329] CNBC, Twitter has suspended more than 1.2 million terrorism-related accounts since late 2015, (2018). Available at: https://www.cnbc.com/2018/04/05/twitter-has-suspended-more-than-1-point-2-million-terrorism-related-accounts.html

[330] The New Arab, Facebook takes action on 1.9m 'terror posts', (2018). Available at: https://www.newarab.com/news/facebook-takes-action-19m-terror-posts

[331] Home Office Interim Code of Practice on Terrorist Content and Activity Online (2020). Available at: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/944036/1704b_ICOP__online_terrorist_content_v.2_11-12-20.pdf

[332] Tech against terrorism :Report: The Threat of Terrorist and Violent Extremist Operated Websites (2022). Available at: https://www.techagainstterrorism.org/research/

[333] Tech against terrorism :Report: The Threat of Terrorist and Violent Extremist Operated Websites (2022).

[334] The Irish Times: Dissident republicans operate ‘broad, unsophisticated online networks’ (2019). Available at: https://www.irishtimes.com/news/crime-and-law/dissident-republicans-operate-broad-unsophisticated-online-networks-1.3869607

[335] Council for Media Services and Reset, The Bratislava Shooting: Report on the role of online platforms (2023). Available at: https://rpms.sk/sites/default/files/2023-03/CMS_RESET_Report.pdf

[336] Mølmen, G and Ravndal, J. Mechanisms of online radicalisation: how the internet affects the radicalisation of extreme-right lone actor terrorists. (2021). Available at: https://www.tandfonline.com/doi/full/10.1080/19434472.2021.1993302

[337] Mølmen, G and Ravndal, J. Mechanisms of online radicalisation: how the internet affects the radicalisation of extreme-right lone actor terrorists. (2021).

[338] The Verge, Facebook says the Christchurch attack live stream was viewed by fewer than 200 people. (2023). Available at: https://www.theverge.com/2019/3/19/18272342/facebook-christchurch-terrorist-attack-views-report-takedown

[339] Regulation to address the dissemination of terrorist content online (European Commission, 2022). Available at: https://home-affairs.ec.europa.eu/policies/internal-security/counter-terrorism-and-radicalisation/prevention-radicalisation/terrorist-content-online_en

[340] Resolution 2617 (2021) (United Nations, 2021). Available at: http://unscr.com/en/resolutions/doc/2617

[341] Interim Code of Practice on Terrorist Content and Activity Online (UK Home Office, 2020). Available at: https://www.gov.uk/government/publications/online-harms-interim-codes-of-practice

[342] Jihadist content targeted on Internet Archive platform (Europol, 2021), Available at: https://www.europol.europa.eu/media-press/newsroom/news/jihadist-content-targeted-internet-archive-platform

[343] Together, we’re tackling online terrorism (Counter Terrorism Policing, 2018). Available at: https://www.counterterrorism.police.uk/together-were-tackling-online-terrorism/

[344] New online tool detects and blocks terrorist propaganda (Sky News, 2018). Available at: https://news.sky.com/story/new-online-tool-detects-and-blocks-terrorist-propaganda-11248450

[345] Macdonald et al, Regulating terrorist content on social media: automation and the rule of law (2019). Available at: https://www.cambridge.org/core/journals/international-journal-of-law-in-context/article/regulating-terrorist-content-on-social-media-automation-and-the-rule-of-law/B54E339425753A66FECD1F592B9783A1

5.10 | Online content associated with child sexual abuse

[346] World Health Organisation, Child Sexual Abuse: A Silent Health Emergency (2004). Available at: https://apps.who.int/iris/bitstream/handle/10665/1878/AFR.RC54.15%20Rev.1.pdf?sequence=1%26isAllowed=y

[347] WeProtect Global Alliance: Global Threat Assessment of child sexual exploitation and abuse online (2021). Available at: https://www.weprotect.org/wp-content/plugins/pdfjs-viewer-shortcode/pdfjs/web/viewer.php?file=https://www.weprotect.org/wp-content/uploads/Global-Threat-Assessment-2021.pdf&attachment_id=143651&dButton=true&pButton=true&oButton=false&sButton=true#zoom=0&pagemode=none&_wpnonce=1fa573dc02

[348] End Violence Against Children, Ruby's Story (2020). Available at: https://www.end-violence.org/articles/rubys-story

[349] NSPCC: Why language matters: why we should never use 'child pornography' and always say child sexual abuse material (2023). Available at: https://learning.nspcc.org.uk/news/why-language-matters/child-sexual-abuse-material

[350] WeProtect Global Alliance The Global Threat Assessment of Child Sexual Exploitation and Abuse Online (2021). Available at: https://www.weprotect.org/wp-content/plugins/pdfjs-viewer-shortcode/pdfjs/web/viewer.php?file=https://www.weprotect.org/wp-content/uploads/Global-Threat-Assessment-2021.pdf&attachment_id=143651&dButton=true&pButton=true&oButton=false&sButton=true#zoom=0&pagemode=none&_wpnonce=4c1339ee30

[351] Wired: The paedophile scandal shows YouTube is broken. Only radical change can fix it (2019). Available at: https://www.wired.co.uk/article/youtube-paedophiles-boycott-algorithm-change

[352] NCMEC, CyberTipline 2022 Report (2022). Available at: https://www.missingkids.org/gethelpnow/cybertipline/cybertiplinedata#:~:text=In%202022%2C%20NCMEC%E2%80%99s%20CyberTipline%20received%2032%20million%20reports,child%20pornography%2C%20make%20up%20the%20largest%20reporting%20category.

[353] 'NCMEC Our 2022 Impact' (2022). Available at: https://www.missingkids.org/ourwork/impact

[354] IWF, Annual Report 2022 (2022). Available at: https://annualreport2022.iwf.org.uk

[355] 'Europe remains global hub for hosting of child sexual abuse material (IWF, 2022). Available at: https://www.iwf.org.uk/news-media/news/europe-remains-global-hub-for-hosting-of-online-child-sexual-abuse-material/#:~:text=In%202021%2041%25%20%28102%2C600%29%20of%20the%20URLs%20were,the%20IWF%20to%20be%20hosted%20in%20the%20Netherlands.

[356] Western People, Hotline ie removed over 14,000 pieces of online child sexual abuse material in 2021 (2022).

[357] Annual Report 2021 (IWF, 2021). Available at: https://annualreport2021.iwf.org.uk/trends/selfgenerated

[358] https://human-rights-channel.coe.int/stop-child-sexual-abuse-in-sport-en.html#:~:text=and%20their%20rights%3F&text=Adults%20must%20break%20the%20silence!&text=1%20in%205-,About%20one%20in%20five%20children%20in%20Europe%20are%20victims%20of,online%20sexual%20extortion%20and%20coercion%E2%80%A6

[359] WeProtect Global Alliance The Global Threat Assessment of Child Sexual Exploitation and Abuse Online (2021). Available at: https://www.weprotect.org/wp-content/plugins/pdfjs-viewer-shortcode/pdfjs/web/viewer.php?file=https://www.weprotect.org/wp-content/uploads/Global-Threat-Assessment-2021.pdf&attachment_id=143651&dButton=true&pButton=true&oButton=false&sButton=true#zoom=0&pagemode=none&_wpnonce=4c1339ee30

[360] The Report of the Independent Inquiry into Child Sexual Abuse Executive Summary (2022). Available at: https://www.iicsa.org.uk/reports-recommendations/publications/inquiry/final-report/executive-summary.html

[361] UK Home Office, The economic and social cost of contact child sexual abuse (2021). Available at: https://www.gov.uk/government/publications/the-economic-and-social-cost-of-contact-child-sexual-abuse/the-economic-and-social-cost-of-contact-child-sexual-abuse

[362] NSPCC, 'Protecting Children from Abuse Online' (2022). Available at: https://learning.nspcc.org.uk/child-abuse-and-neglect/child-sexual-abuse

[363] Hailes, Yu, Danese, Fazel, 'Long-term outcomes of childhood sexual abuse: an umbrella review (2019). Available at: https://www.thelancet.com/journals/lanpsy/article/PIIS2215-0366(19)30286-X/fulltext

[364] Australian Institute of Criminology, Production and distribution of child sexual abuse material by parental figures (2021) Accessed from: https://www.aic.gov.au/sites/default/files/2021-02/ti616_production_and_distribution_of_child_sexual_abuse_material_by_parental_figures.pdf 09/03/2021

[365] Catch 22, 16 Days of Action: The impact of child sexual abuse on victims and survivors (2022). Available at: https://www.catch-22.org.uk/resources/the-impact-of-child-sexual-abuse-on-victims-and-survivors/

[366] See 144 to 148.

[367] European Commission, Communication from the Commission to the European Parliament, the council, the European economic and social Committee and the Committee of the regions: EU Strategy for a more effective fight against Child Sexual Abuse (2022). Available at: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52020DC0607

[368] United Nations Office on Drugs and Crime, Study on the effects of new information technologies on the abuse and exploitation of children (2015) Accessed from: https://www.unodc.org/documents/Cybercrime/Study_on_the_Effects.pdf22/04/2021

[369] Canadian Centre for Child Protection, 'Project Arachnid: Online Availability of Child Sexual Abuse Material' (2021). Available at: https://protectchildren.ca/pdfs/C3P_ProjectArachnidReport_en.pdf

[370] Australian E-Safety Commissioner, 'Basic Online Safety Expectations (2022). Available at: https://apo.org.au/sites/default/files/resource-files/2022-12/apo-nid321193.pdf

[371] Yubo, FAQ. Available at: https://www.yubo.live/faq

[372] WeProtect Global Alliance The Global Threat Assessment of Child Sexual Exploitation and Abuse Online (2021). Available at: https://www.weprotect.org/wp-content/plugins/pdfjs-viewer-shortcode/pdfjs/web/viewer.php?file=https://www.weprotect.org/wp-content/uploads/Global-Threat-Assessment-2021.pdf&attachment_id=143651&dButton=true&pButton=true&oButton=false&sButton=true#zoom=0&pagemode=none&_wpnonce=4c1339ee30

[373] Ionos,'Upload filters: a danger to free internet content?' (2019). Available at: https://www.ionos.com/digitalguide/websites/digital-law/upload-filters

[374] Meta, Preventing Child Exploitation on our Apps (2021). Available at: https://about.fb.com/news/2021/02/preventing-child-exploitation-on-our-apps/

[375] The Washington Post. AI-generated child sex images spawn new nightmare for the web. (2023). Available at: https://www.washingtonpost.com/technology/2023/06/19/artificial-intelligence-child-sex-abuse-images/

[376] BBC, Illegal trade in AI child sex abuse images exposed. (2023) Available at: https://www.bbc.co.uk/news/uk-65932372

[377] BBC, Illegal trade in AI child sex abuse images exposed. (2023)

[378] Google, https://protectingchildren.google/#fighting-abuse-on-our-own-platform-and-services . Thorn, https://www.thorn.org/blog/how-safers-detection-technology-stops-the-spread-of-csam/

[379] Yubo, FAQ. Available at: https://www.yubo.live/faq

[380] SafeToNet, SafeToWatch (2023). Available at: https://safetonet.com/safetowatch/

[381] OFCOM, Overview of Perceptual Hashing Technology, (2022.) Available at: https://www.ofcom.org.uk/__data/assets/pdf_file/0036/247977/Perceptual-hashing-technology.pdf

[382] IWF, 'Self-generated' child sexual abuse prevention campaign' (2021). Available at: https://www.iwf.org.uk/about-us/our-campaigns/talk-and-gurls-out-loud-self-generated-child-sexual-abuse-prevention-campaign/

[383] This can be found in the Online Safety and Media Regulation Act 2022, Schedule 3, Section 38, 39 and 40.

5.11 | Online content by which a person's behaviour constitutes harassment or harmful communication

[384] Houses of the Oireachtas, Harassment, Harmful Communications and Related Offences Act (2020). Available at: https://data.oireachtas.ie/ie/oireachtas/act/2020/32/eng/enacted/a3220.pdf

[385] Houses of the Oireachtas, Non-fatal Offences Against the Person Act (1997). Available at: https://www.irishstatutebook.ie/eli/1997/act/26/enacted/en/print.html#:~:text=or%20to%20both.-,Harassment.,be%20guilty%20of%20an%20offence.

[386] Women's Aid, One in five (2020).

[387] Cybersmile, Stop Cyberbullying Day Survey (2017). Available at: https://www.cybersmile.org/wp-content/uploads/Stop-Cyberbullying-Day-Survey-2017.pdf

[388] Office of the attorney General, Equality Act 2004 (2004). Available at: https://www.irishstatutebook.ie/eli/2004/act/24/enacted/en/print#sec8

[389] This can be found in the Online Safety and Media Regulation Act 2022 in Schedule 3 Sections 13, 38, 39, 40 and 41.

[390] Universities UK, Tackling online harassment and promoting online welfare (2019) Available at: https://www.universitiesuk.ac.uk/sites/default/files/field/downloads/2021-07/tackling-online-harassment.pdf

[391] FRA (EU agency for fundamental rights), Fundamental rights survey (2019). Available at: https://fra.europa.eu/en/data-and-maps/2021/frs?mdq1=theme&mdq2=982

[392] Victims Commissioner, The Impact of Online Abuse: Hearing the Victim's Voice (2022). Available at: https://cloud-platform-e218f50a4812967ba1215eaecede923f.s3.amazonaws.com/uploads/sites/6/2022/05/Hearing-the-Victims-Voice.pdf

[393] European Union, Fundamental Rights Survey. (2020) Available at: https://fra.europa.eu/en/data-and-maps/2021/frs?mdq1=theme&mdq2=982

[394] Plan International, Free to be online? (2020). Available at: https://plan-international.org/uploads/2022/02/sotwgr2020-commsreport-en-2.pdf

[395] Women's Aid, One in five (2020).

[396] Plan International, Free to be online? (2020).

[397] Women's Aid, One in five (2020).

[398] European Commission, Project deSHAME: Young people's experiences of online sexual harassment. (2018)

[399] European Commission, Project deSHAME: Young people's experiences of online sexual harassment. (2018)

[400] Dublin City University, The Gendered Experiences of Image-based Sexual Abuse: State of the Research and Evidence-based Recommendations (2022).

[401] Childnet, Project deShame Report (2017).

[402] European Commission, Project deSHAME: Young people's experiences of online sexual harassment. (2018)

[403] Cybersmile, Stop Cyberbullying Day Survey (2017).

[404] Revenge Porn Helpline (RPH), Revenge Porn Helpline Report (2022). Available at: https://revengepornhelpline.org.uk/assets/documents/rph-report-2022.pdf?_=1681885542

[405] CNN, Tiziana Cantone's family calls for justice after suicide over sex tape (2016). Available at: https://edition.cnn.com/2016/09/16/europe/tiziana-cantone-sex-tape-suicide/index.html

[406] Revenge Porn Helpline (RPH), Revenge Porn Helpline Report (2022).

[407] Women's Aid, One in five (2020).

[408] Childnet, Project deShame Report (2017). Available at: https://www.childnet.com/what-we-do/our-projects/project-deshame/research/

[409] Plan International, Free to be online? (2020).

[410] Women's Aid, One in five (2020).

[411] Victims Commissioner, The Impact of Online Abuse: Hearing the Victim's Voice (2022).

[412] Dublin City University, The Gendered Experiences of Image-based Sexual Abuse: State of the Research and Evidence-based Recommendations (2022).

[413] European Commission, Project deSHAME: Young people's experiences of online sexual harassment. (2018)

[414] Revenge Porn Helpline (RPH), Revenge Porn Helpline Report (2022).

[415] Childnet, Project deShame Report (2017).

[416] Cybersmile, Stop Cyberbullying Day Survey (2017).

[417] Plan International, Free to be online? (2020).

[418] House of Commons Petitions Committee, Tackling Online Abuse (2022).

[419] Refuge, Marked as Unsafe (2022) Available at: https://refuge.org.uk/wp-content/uploads/2022/11/Marked-as-Unsafe-FINAL-November-2022.pdf?utm_source=Social+Media&utm_medium=Twitter&utm_campaign=MAUS

[420] Women's Aid, One in five (2020).

[421] Women's Aid, One in five (2020).

[422] Plan International, Free to be online? (2020).

[423] Dublin City University, The Gendered Experiences of Image-based Sexual Abuse: State of the Research and Evidence-based Recommendations (2022).

[424] Women's Aid, One in five (2020).

[425] European Commission, Project deSHAME: Young people's experiences of online sexual harassment. (2018)

[426] European Commission, Project deSHAME: Young people's experiences of online sexual harassment. (2018)

[427] Dublin City University, The Gendered Experiences of Image-based Sexual Abuse: State of the Research and Evidence-based Recommendations (2022).

[428] Dublin City University, Young People's Experiences of Sexual and Gender-based Harassment and Abuse During the Covid-19 Pandemic in Ireland: Incidence, Intervention and Recommendations (2022).

[429] Women's Aid, One in five (2020).

[430] United Nations UNITAD, Cyber-harassment: self-protection tips (2023). Available at: https://www.unitad.un.org/content/cyber-harassment-self-protection-tips

[431] Meta, Facebook community standards (2023). Available at: https://transparency.fb.com/en-gb/policies/community-standards/

[432] European Commission, Project deSHAME: Young people's experiences of online sexual harassment. (2018)

[433] Dublin City University, Young People's Experiences of Sexual and Gender-based Harassment and Abuse During the Covid-19 Pandemic in Ireland: Incidence, Intervention and Recommendations (2022).

[434] SWGfl, Minerva – Record Online Abuse (2023). Available at: https://swgfl.org.uk/helplines/minerva/

[435] House of Commons Petitions Committee, Tackling Online Abuse (2022). Available at: https://committees.parliament.uk/publications/8669/documents/89002/default/

[436] Jigsaw, Harassment Manager (2023). Available at: https://jigsaw.google.com/harassment-manager/

[437] Audiovisual media services directive 2018

[438] IMD Business School: Digital ecosystems (2019). Available at https://www.imd.org/reflections/digital-ecosystems/

5.12 | Online content associated audio-visual commercial communications

[439] Kozinets, R., How social media is helping Big Tobacco hook a new generation of smokers. (2019). Available at: https://theconversation.com/how-social-media-is-helping-big-tobacco-hook-a-new-generation-of-smokers-112911

[440] Audiovisual media services directive 2018

[441] European Audiovisual Observatory, Mapping report on the rules applicable to video-sharing platforms: Focus on commercial communications (2022) Available at: https://rm.coe.int/mapping-on-video-sharing-platforms-2022-focus-on-cc/1680aa1b15

[442] European Audiovisual Observatory, Mapping report on the rules applicable to video-sharing platforms: Focus on commercial communications (2022)

[443] Included in the AVMSD but not the OSMR

[444] Included in the AVMSD but not the OSMR

[445] Included in the OSMR but not the AVMSD

[446] Ofcom, Video-sharing platform guidance (2021) Available at: https://www.ofcom.org.uk/__data/assets/pdf_file/0014/229010/vsp-guidance-control-of-advertising.pdf

[447] Ofcom, Video-sharing platform guidance (2021)

[448] Ofcom, Video-sharing platform guidance (2021)

[449] Competition and Consumer Protection Commission, Online behaviour: Influencer marketing (2022) Available at: https://www.ccpc.ie/business/wp-content/uploads/sites/3/2022/12/2022.12.12-172837-CCPC-Influencer-marketing-report.pdf

[450] Advertising Standards Authority for Ireland, Artificial Intelligence tools to be implemented by Advertising Standards Authority for Ireland to proactively identify social media posts by influencers and breaches of the advertising code (2022) Available at: https://www.asai.ie/press-releases/artificial-intelligence-tools-to-be-implemented-by-advertising-standards-authority-for-ireland-to-proactively-identify-social-media-posts-by-influencers-and-breaches-of-the-advertising-code/

[451] Ofcom, Online Nation 2022 Report (2022) Available at: https://www.ofcom.org.uk/__data/assets/pdf_file/0023/238361/online-nation-2022-report.pdf

[452] YONDER, User experience of potential online harms within video sharing platforms (2021)

[453] Mardon, R. YouTube's child viewers may struggle to recognise adverts in videos from 'virtual play dates' (2019). Available at: https://blogs.lse.ac.uk/parenting4digitalfuture/2019/09/25/youtubes-child-viewers-may-struggle-to-recognise-adverts/

[454] Jigsaw (2020). Internet users' concerns about and potential experience of potential online harms. Ofcom. Available at: https://www.ofcom.org.uk/__data/assets/pdf_file/0024/196413/concerns-and-experiences-online-harms-2020-chart-pack.pdf

[455] Ofcom, Children's Media Lives (Ongoing) Available at: https://www.ofcom.org.uk/research-and-data/media-literacy-research/childrens/childrens-media-lives

[456] Safe food advocacy Europe, European Commission releases study on children exposure to marketing of high in fat, salt or sugar (HFSS) foods (2021). Available at: https://www.safefoodadvocacy.eu/european-commission-releases-study-on-children-exposure-to-marketing-of-high-in-fat-salt-or-sugar-hfss-foods/

[457] Keller A., Social media and alcohol (2020) Available at: https://www.drugrehab.com/addiction/alcohol/influence-of-social-media/

[458] Griffin et. al, All night long: Social media marketing to young people by alcohol brands and venues (2018) Available at: https://alcoholchange.org.uk/publication/all-night-long-social-media-marketing-to-young-people-by-alcohol-brands-and-venues

[459] Hendrilks et al, Picture Me Drinking: Alcohol-Related Posts by Instagram Influencers Popular Among Adolescents and Young Adults (2019). Available at: https://pubmed.ncbi.nlm.nih.gov/32038379/

[460] The Guardian, Instagram influencers advertising nicotine products to young people, charity warns, (2023). Available at:https://www.theguardian.com/business/2023/jun/12/instagram-influencers-advertising-nicotine-products-to-young-people-charity-warns

[461] Kozinets, R., How social media is helping Big Tobacco hook a new generation of smokers. (2019).

[462] Department of Health and Social Care: Introducing a total online advertising restriction for products high in fat, sugar and salt (HFSS) – (2021) https://www.gov.uk/government/consultations/total-restriction-of-online-advertising-for-products-high-in-fat-sugar-and-salt-hfss/introducing-a-total-online-advertising-restriction-for-products-high-in-fat-sugar-and-salt-hfss#children39s-media-habits-and-hfss-advertising-online

[463] ISPCC, What is Neuromarketing and how does it affect our children? (2023) Available at: https://www.ispcc.ie/what-is-neuromarketing-and-how-does-it-affect-our-children/

[464] Abbas, AA. et al., The Impact of Neuromarketing Advertising on Children: Intended and unintended effects (2019) Available at: https://knepublishing.com/index.php/Kne-Social/article/view/5187/10293

[465] ISPCC, What is Neuromarketing and how does it affect our children? (2023)

[466] Abbas, AA. et al., The Impact of Neuromarketing Advertising on Children: Intended and unintended effects (2019)

[467] Competition and Consumer Protection Commission, Online behaviour: Influencer marketing (2022)

[468] Davidson et al, Research on Protection of Minors: A literature review and interconnected frameworks. Implications for VSP regulation and beyond (2021)

[469] Facebook ad targeting could be used to incite sectarian violence, Global Witness Organisation (2021). Available at: https://www.globalwitness.org/en/press-releases/facebook-ad-targeting-could-be-used-incite-sectarian-violence/

[470] The Irish Times, Hateful ads submitted to Facebook, TikTok and YouTube approved by platforms (2023). Available at: https://www.irishtimes.com/technology/big-tech/2023/02/23/hateful-ads-submitted-to-facebook-tiktok-and-youtube-approved-by-platforms/

[471] Ofcom, The regulation of advertising on video-sharing platforms (2021) Available at: https://www.ofcom.org.uk/__data/assets/pdf_file/0022/229009/vsp-advertising-statement.pdf

[472] Ofcom, The regulation of advertising on video-sharing platforms (2021) Available at: https://www.ofcom.org.uk/__data/assets/pdf_file/0022/229009/vsp-advertising-statement.pdf

[473] Ofcom, The regulation of advertising on video-sharing platforms (2021)

[474] Ofcom, The regulation of advertising on video-sharing platforms (2021)

[475] YONDER, User experience of potential online harms within video sharing platforms (2021)

[476] YONDER, User experience of potential online harms within video sharing platforms (2021)

[477] YONDER, User experience of potential online harms within video sharing platforms (2021)

[478] Yonder, Project 2 Qualitative – Safety measures on video-sharing platforms survey (2020) Available at: https://www.ofcom.org.uk/__data/assets/pdf_file/0021/216516/safety-measures-vsp-survey-2021-qual.pdf

[479] Competition and Consumer Protection Commission, Online behaviour: Influencer marketing (2022)

[480] YONDER, User experience of potential online harms within video sharing platforms (2021)

[481] YONDER, Project 3: VSP Content creators and community standards (2021) Available at: https://www.ofcom.org.uk/__data/assets/pdf_file/0024/216519/content-creators-community-standards.pdf

[482] SuperAwesome, 7 best practices for kid-safe influencer marketing (2020) Available at: https://www.superawesome.com/sacreators-best-practices-e-book/

[483] Advertising Standards Authority, Influencers' guide to making clear that ads are ads (2023) Available at: https://www.asa.org.uk/resource/influencers-guide.html

[484] Competition and Consumer Protection Commission, Online behaviour: Influencer marketing (2022)

[485] Davidson et al, Research on Protection of Minors: A literature review and interconnected frameworks. Implications for VSP regulation and beyond (2021)

6 | General response measures

[486] Council of Europe, Digital Citizenship Education. Available at: https://www.coe.int/en/web/digital-citizenship-education/the-concept

[487] Lawfare, Time for Transparency From Digital Platforms, But What Does That Really Mean? (2022). Available at:https://www.lawfareblog.com/time-transparency-digital-platforms-what-does-really-mean

7 | Outstanding matters

[488] Similar approaches can be found in the EU Digital Services Act and the UK Online Safety Bill.

[489] Meta, 2023: ‘Transparency Center’. Available at: https://transparency.fb.com/en-gb/

[490] Twitter, 2023: ‘Transparency’. Available at: https://transparency.twitter.com/

[491] E-Safety Commissioner, 2022: ‘Basic Online Safety Expectations: summary of industry responses to the first mandatory transparency notices’. Available at: https://www.esafety.gov.au/sites/default/files/2022-12/BOSE%20transparency%20report%20Dec%202022.pdf

[492] The Santa Clara Principles on Transparency and Accountability in Content Moderation call for companies to clearly outline what controls users have access to which enable them to manage how their content is curated using algorithmic systems, and what impact these controls have over a user’s online experience.

[493] OFCOM, Transparency in the regulation of online safety, (2021). Available at: https://www.ofcom.org.uk/__data/assets/pdf_file/0020/220448/transparency-in-online-safety.pdf

[494] Brandies, ‘Chapter V- What Publicity Can Do’ (1914). Available at: OTHER PEOPLE'S MONEY - CHAPTER V — Louis D. Brandeis School of Law Library (louisville.edu)

[495] European Parliament, ‘Corporate due diligence and corporate accountability: European Parliament resolution of 10 March 2021 with recommendations to the Commission on corporate due diligence and corporate accountability’ (2021). Available at: https://www.europarl.europa.eu/doceo/document/TA-9-2021-0073_EN.pdf

[496] Birkinshaw, ‘Transparency as a Human Right’ [Abstract] (2006) Available at: https://britishacademy.universitypressscholarship.com/view/10.5871/bacad/9780197263839.001.0001/upso-9780197263839-chapter-3

[497] Global Network Initiative: ‘The GNI Principles’ https://globalnetworkinitiative.org/gni-principles

[498] Children’s Research Network, ‘Do children have rights in the online world and can they be enforced? An analysis of the Council of Europe’s ‘Guidelines on Children’s Rights in the Digital Environment’ from an Irish perspective.’ (2020). Available at: https://www.childrensresearchnetwork.org/knowledge/resources/article-child-rights-online-world

[499] Møllgaard & Overgaard, ‘Market transparency competition policy’ (2021). Available at: https://www.econstor.eu/bitstream/10419/208446/1/cbs-wp2001-06.pdf

[500] Digital Services Act: agreement for a transparent and safe online environment (2022). Available at: https://www.europarl.europa.eu/news/en/press-room/20220412IPR27111/digital-services-act-agreement-for-a-transparent-and-safe-online-environment

[501] Department for Digital, Culture, Media & Sport, The government report on transparency reporting in relation to online harms (2020).

[502] The Justice Collaboratory, Report of the Facebook data transparency advisory group (2019). Available at: https://law.yale.edu/sites/default/files/area/center/justice/document/dtag_report_5.22.2019.pdf

[503] FSA: 'Discussion Paper DP13/1 Transparency' https://www.fca.org.uk/publication/discussion/fsa-dp13-01.pdf

[504] Transatlantic Working Group, 2019: 'An analysis of Germany's NetzDG Law'. Available at: https://www.ivir.nl/publicaties/download/NetzDG_Tworek_Leerssen_April_2019.pdf

[505] Urman, A., Makhortykh, M., How transparent are transparency reports? Comparative analysis of transparency reporting across online platforms (2023). Available at: https://www.sciencedirect.com/science/article/pii/S0308596122001793

[506] OFCOM, Transparency in the regulation of online safety, (2021).

[507] OFCOM, Transparency in the regulation of online safety, (2021).

[508] Department for Digital, Culture, Media & Sport, The government report on transparency reporting in relation to online harms (2020). Available at: https://www.gov.uk/government/consultations/online-harms-white-paper/outcome/government-transparency-report#:~:text=Transparency%20reports%20will%20provide%20users,analysis%20from%20the%20company%20reports.

[509] Gaventa & McGee, 'The Impact of Transparency and Accountability Initiatives' (2013). Available at: https://assets.publishing.service.gov.uk/media/57a08aabed915d622c00084b/60827_DPRGaventaMcGee_Preprint.pdf

[510] Aspen Institute. Final Report: Commission on Information Disorder. (2021) Available at: https://www.aspeninstitute.org/wp-content/uploads/2021/11/Aspen-Institute_Commission-on-Information-Disorder_Final-Report.pdf

[511] Council of the European Union. EU Introduces Transparency Obligations for Online Platforms. (2021) Available at: https://www.consilium.europa.eu/en/press/press-releases/2019/06/14/eu-introduces-transparency-obligations-for-online-platforms/

[512] Dawes, Dame M. In News We Trust: Keeping Faith in the Future of Media. 2021. Available at: https://www.ofcom.org.uk/news-centre/2021/keeping-faith-in-future-of-media

[513] SSRN Scholarly Paper; Integrity Institute. Metric & Transparency: Data and Datasets to Track Harms, Design, and Process on Social Media Platforms (2021) Available at: https://static1.squarespace.com/static/614cbb3258c5c87026497577/t/617834d31bcf2c5ac4c07494/1635267795944/Metrics+and+Transparency+-+Summary+%28EXTERNAL%29.pdf

[514] LinkedIn, 'What are the benefits and risks of using AI or automated tools for content moderation of live media?' (2023). Available at: https://www.linkedin.com/advice/0/what-benefits-risks-using-ai-automated-tools-content#:~:text=AI%20or%20automated%20tools%20can%20offer%20some%20advantages,harmful%20content%20based%20on%20predefined%20rules%20or%20criteria.

[515] Medium, 'News feeds, old content: a brief history of algorithmically curated feeds on Facebook and Twitter' (2019). Available at: https://medium.com/@annawchung/news-feeds-old-content-a-brief-history-of-algorithmically-curated-feeds-on-facebook-and-twitter-85b5e5d8e30a.

[516] TSPA.org, 'What is content moderation?' (2023). Available at: https://www.tspa.org/curriculum/ts-fundamentals/content-moderation-and-operations/what-is-content-moderation/

[517] Information Commissioner's Office, 'Rights related to automated decision-making, including profiling' (2023). Available at: https://ico.org.uk/for-organisations/uk-gdpr-guidance-and-resources/individual-rights/individual-rights/rights-related-to-automated-decision-making-including-profiling/#ib2

[518] Statista, 'Social media statistics 2023: top networks by the numbers' (2023). Available at:

[519] Forbes, 'The growing role of AI in content moderation' (2022). Available at: https://dustinstout.com/social-media-statistics/https://www.forbes.com/sites/forbestechcouncil/2022/06/14/the-growing-role-of-ai-in-content-moderation/?sh=fc425f74a178

[520] Tarleton Gillespie, 'Content moderation, AI, and the question of scale' (2020). Available at: https://journals.sagepub.com/doi/10.1177/2053951720943234

[521] Department for Digital, Culture, Media, and Sport, 'Online Harms Feasibility Study' (2021). Available at: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1124858/DCMS_Online_Harm_Feasibility_study_v2.pdf

[522] Tarleton Gillespie, 'Content moderation, AI, and the question of scale' (2020). Available at: https://journals.sagepub.com/doi/10.1177/2053951720943234

[523] Tarleton Gillespie, 'Content moderation, AI, and the question of scale' (2020).

[524] Davidson et al, Racial Bias in Hate Speech and Abusive Language Detection Datasets, (2019). Available at: https://arxiv.org/pdf/1905.12516.pdf

[525] Tarleton Gillespie, 'Content moderation, AI, and the question of scale' (2020).

[526] Buolamwini J, Gebru T (2018) Gender shades: Intersectional accuracy disparities in commercial gender classification. Proceedings of Machine Learning Research 81: 1–15

[527] Eugenia Siapera, 'AI Content Moderation, Racism and (de)Coloniality' (2022). Available at: https://link.springer.com/article/10.1007/s42380-021-00105-7

[528] ComputerWeekly, 'The one problem with AI content moderation? It doesn't work' (2023). Available at: https://www.computerweekly.com/feature/The-one-problem-with-AI-content-moderation-It-doesnt-work

[529] National Centre for Missing and Exploited Children, 'Child Sexual Abuse Material' (2023). Available at: https://www.computerweekly.com/feature/The-one-problem-with-AI-content-moderation-It-doesnt-work

[530] Make Use Of, 'How do social media feed algorithms work?' (2021). Available at: https://www.makeuseof.com/how-social-media-algorithms-work/

[531] The Irish Times, 'Irish people spend 4.5 hours on their smartphones daily' (2019). Available at: https://www.statista.com/statistics/719858/average-daily-internet-and-social-media-use-in-ireland-by-device/

[532] Statista, 'Daily time spent with selected media among adults in Ireland in 3rd quarter 2022.' (2022). Available at: https://www.statista.com/statistics/719858/average-daily-internet-and-social-media-use-in-ireland-by-device/

[533] The Irish News, 'Primary school children miss out on sleep as they scroll through social media' (2022). Available at: https://www.irishnews.com/magazine/science/2022/09/14/news/_primary_school_children_miss_out_on_sleep_as_they_scroll_through_social_media_-2828649/

[534] British Medical Journal, 'Is excessive use of social media an addition?' (2019). Available at: https://www.bmj.com/content/365/bmj.l2171

[535] Harvard Business Review, 'The Psychology of your Scrolling Addiction' (2022). Available at: https://hbr.org/2022/01/the-psychology-of-your-scrolling-addiction

[536] Med-Phoenix Journal of National Medical College, 'Excess screen time: impact on childhood development and management' (2021). Available at: https://www.researchgate.net/publication/353367437_Excess_Screen_Time_-_Impact_on_Childhood_Development_and_Management_A_Review

[537] Wired, 'The paedophile scandal shows YouTube is broken. Only radical change can fix it' (2019). Available at: https://www.wired.co.uk/article/youtube-paedophiles-boycott-algorithm-change

[538] Center for Countering Digital Hate, Deadly by Design (2022)

[539] Center for Countering Digital Hate, Deadly by Design (2022)

[540] NY Times, 'Rabbit Hole' (2023). Available at: https://www.nytimes.com/column/rabbit-hole

[541] GOV.UK, 'The benefits and harms of algorithms: a shared perspective from the four digital regulators' (2022). Available at: https://www.gov.uk/government/publications/findings-from-the-drcf-algorithmic-processing-workstream-spring-2022/the-benefits-and-harms-of-algorithms-a-shared-perspective-from-the-four-digital-regulators#current-and-potential-harms-of-algorithmic-processing

[542] MIT Technology Review, 'The Facebook whistleblower says its algorithms are dangerous. Here's why.' (2021). Available at: https://www.technologyreview.com/2021/10/05/1036519/facebook-whistleblower-frances-haugen-algorithms/

[543] GOV.UK, 'The benefits and harms of algorithms: a shared perspective from the four digital regulators' (2022). Available at: https://www.gov.uk/government/publications/findings-from-the-drcf-algorithmic-processing-workstream-spring-2022/the-benefits-and-harms-of-algorithms-a-shared-perspective-from-the-four-digital-regulators#current-and-potential-harms-of-algorithmic-processing

[544] Government of Ireland, 'Harnessing Digital: The Digital Ireland Framework' (2022). Available at: https://www.gov.ie/pdf/?file=https://assets.gov.ie/241714/bedc64c6-baaf-4100-9255-02f5e07dd3f9.pdf#page=null

[545] European Commission, 'European Declaration on Digital Rights and Principles for the Digital Decade' (2022). Available at: https://ec.europa.eu/newsroom/dae/redirection/document/82703

[546] United Nations, 'Draft recommendation on the ethics of artificial intelligence, online' (2021). Available at: https://unesdoc.unesco.org/search/N-EXPLORE-4407ecc2-5700-4c44-b729-a7473e963814

[547] SantaClaraPrinciples.org, 'The Santa Clara Principles' (2023). Available at: https://www.bbc.co.uk/news/technology-64813981https://santaclaraprinciples.org/

[548] BBC News, 'TikTok sets 60-minute daily screen time limit for under-18s.' (2023). Available at:

[549] Parliament UK, 'Online Safety Bill' (2023). Available at: https://bills.parliament.uk/publications/49376/documents/2822

[550] The UK's Online Safety Bill defines U2U services as an internet service where content that is generated directly on the service by a user or uploaded to or shared on the service by a user, may be encountered by another user, or users, of the service.

[551] PATTRN.AI, Evaluation recommender systems in relation to the dissemination of illegal and harmful content in the UK (2023) https://www.ofcom.org.uk/__data/assets/pdf_file/0029/263765/Pattrn_Anayltics_Intelligence_Final_Report.pdf

[552] Politico, 'EU lawmakers vote to ban online ads targeting children amid broader tech crackdown' (2021). Available at: https://www.politico.eu/article/eu-lawmaker-rule-out-online-ads-target-children/.

[553] Meta, Meta Earnings Presentation Q4 2022 (2022). Available at: https://s21.q4cdn.com/399680738/files/doc_financials/2022/q4/Earnings-Presentation-Q4-2022.pdf

[554] Yubo, How many people use Yubo?. Available at: https://support.yubo.live/hc/en-us/articles/5303664223762-How-many-people-use-Yubo-

[555] Ofcom, Ofcom's first year of video-sharing platform regulation (2022). Available at: https://www.ofcom.org.uk/__data/assets/pdf_file/0032/245579/2022-vsp-report.pdf

[556] DCMS, Impact Assessment on the Online Safety Bill (2021). Available at: https://publications.parliament.uk/pa/bills/cbill/58-02/0285/onlineimpact.pdf

[557] DCMS, Impact Assessment on the Online Safety Bill (2021)

[558] DCMS, Impact Assessment on the Online Safety Bill (2021).

[559] European Central Bank, Proportionality in banking supervision - Panel intervention by Pentti Hakkarainen (2019). Available at: https://www.bankingsupervision.europa.eu/press/speeches/date/2019/html/ssm.sp190509~7b20eedbe7.en.html

[560] Bank for International Settlements, Speech by Mr Fernando Restoy, Chairman, Financial Stability Institute, Bank for International Settlements at the Westminster Business Forum (2018). Available at: https://www.bis.org/speeches/sp180704b.htm

[561] European Central Bank, Is small beautiful? Supervision, regulation and the size of banks. A statement by Sabine Lautenschläger (2017). Available at: https://www.ecb.europa.eu/press/key/date/2017/html/ecb.sp171014.en.html

[562] European Law, Regulation (EC) No 178/2002 of the European Parliament and the Council of 28 January 2002 laying down the general principle and requirements of food law (2002). Available at: https://eur-lex.europa.eu/legal-content/EN/ALL/?uri=CELEX:32002R0178

[563] Meta Earnings Presentation Q3 2022, (2022). Available at: https://s21.q4cdn.com/399680738/files/doc_financials/2022/q3/Q3-2022_Earnings-Presentation.pdf

[564] Time, We Can Regulate Social Media Without Censorship. Here's How (July 2022). Available at:https://time.com/6199565/regulate-social-media-platform-reduce-risks/

[565] Coventry University, Postdigital Intimacies for Online Safety (May 2023). Available at: https://issuu.com/postdigitalintimacies/docs/online_safety_bill_report_final

[566] DCMS, Impact Assessment on the Online Safety Bill (2021). Available at: https://publications.parliament.uk/pa/bills/cbill/58-02/0285/onlineimpact.pdf

[567] Ofcom, Video Sharing Platforms: Ofcom's Plan and Approach (2021). Available at: https://www.ofcom.org.uk/__data/assets/pdf_file/0016/226303/vsp-plan-approach.pdf

[568] European Central Bank, Regulation, proportionality and the sustainability of banking – Speech by Andrea Enria, Chair of the Supervisory Board of the ECB, at the Retail Banking Conference (2019). Available at: https://www.bankingsupervision.europa.eu/press/speeches/date/2019/html/ssm.sp191121_1~a65cdec01d.en.html

[569] OECD, The SME Test: Taking SMEs and entrepreneurs into account when regulating (2022). Available at: https://one.oecd.org/document/GOV/RPC%282021%2921/FINAL/en/pdf

[570] Harling, A., Henesy, D., Simmance, E. Transparency Reporting: The UK Regulatory Perspective (2023). Available at: https://tsjournal.org/index.php/jots/article/view/108/43

[571] Revealing Reality, How people are harmed online: Testing a model from a user perspective (2022) Available at: https://www.ofcom.org.uk/__data/assets/pdf_file/0023/244238/How-people-are-harmed-online-testing-a-model-from-a-user-perspective.pdf

[572] NCMEC, 2023. Available at: https://www.missingkids.org/

[573] INHOPE, 'INHOPE Annual Report'. (2022). Available at: https://inhope.org/EN/articles/inhope-annual-report-2022

[574] IWF, 'Annual Report 2022'. (2023) Available at: https://annualreport2022.iwf.org.uk/

[575] European Commission, 'Five years of Project Arachnid'. (2022). Available at: https://home-affairs.ec.europa.eu/news/five-years-project-arachnid-2022-01-17_en

[576] Meta, 'Processing under EU Regulation 2021/1232' (2022). Available at: https://transparency.fb.com/sr/eu-csam-derogation-report-2023/

[577] WeProtect Global Alliance, 'Global Threat Assessment of child exploitation and abuse online', (2021). Available at: https://www.weprotect.org/wp-content/plugins/pdfjs-viewer-shortcode/pdfjs/web/viewer.php?file=https://www.weprotect.org/wp-content/uploads/Global-Threat-Assessment-2021.pdf&attachment_id=143651&dButton=true&pButton=true&oButton=false&sButton=true#zoom=0&pagemode=none&_wpnonce=1f0a20b309

[578] Meta, Preventing child exploitation on our apps (2021).

[579] Draft Online Safety Bill, 2021. Available at: https://www.gov.uk/government/publications/draft-online-safety-bill

[580] The Guardian, 'Pornhub removes millions of videos after investigation finds child abuse content.' (2020). Available at:https://www.theguardian.com/technology/2020/dec/14/pornhub-purge-removes-unverified-videos-investigation-child-abuse?ref=hackernoon.com

[581] End Violence, ‘Disrupting Harm: Conversations with young survivors’ (2023). Available at: https://www.end-violence.org/sites/default/files/2023-01/Disrupting%20Harm-Conversations%20with%20young%20survivors%20about%20online%20child%20sexual%20exploitation%20and%20abuse.pdf

[582] SCMP, ‘China’s porn censors shut down 12,000 websites in the first half of 2020’ (2020). Available at: https://www.scmp.com/abacus/news-bites/article/3092512/chinas-porn-censors-shut-down-12000-websites-first-half-2020

[583] ECPAT, 2023: ‘Boys in Morocco are not being recognised as victims of sexual exploitation’. Available at: https://ecpat.org/story/morocco-boys/

[584] WeProtect Global Alliance, ‘Intelligence briefing: The sexual exploitation and abuse of deaf and disabled children online’ (2021). Available at: https://www.weprotect.org/wp-content/uploads/Intelligence-briefing-2021-The-sexual-exploitation-and-abuse-of-disabled-children.pdf

[585] Children’s Commissioner, ‘Young People and Pornography’ (2023). Available at: https://assets.childrenscommissioner.gov.uk/wpuploads/2023/05/CCO-Pornography-and-Young-People.pdf

[586] Department for Digital, Culture, Media, and Sport. Online Harms Feasibility Study (2021).

[587] Ofcom, Just one in six young people flag harmful content online. (2022) Available at: https://www.ofcom.org.uk/news-centre/2022/one-in-six-young-people-flag-harmful-content-online

[588] NACOS: Report of a National Survey of Children, their Parents and Adults regarding Online Safety (2021)

[589] Charter of Fundamental Rights of the European Union (2000). Available at: https://www.europarl.europa.eu/charter/pdf/text_en.pdf

[590] CitizensInformation.ie, Charter of Fundamental Rights (2023). Available at: https://www.citizensinformation.ie/en/government-in-ireland/european-government/eu-law/charter-of-fundamental-rights/#:~:text=The%20Charter%20of%20Fundamental%20Rights,with%20the%20Treaty%20of%20Lisbon.

[591] https://www.irishstatutebook.ie/eli/cons/en/html#article40

[592] European Commission, Questions and Answers: Digital Services Act* (2023). Available at: https://ec.europa.eu/commission/presscorner/detail/en/QANDA_20_2348

[593] Council of Europe, Human Rights for Internet Users. Available at: https://www.coe.int/en/web/freedom-expression/guide-to-human-rights-for-internet-users

[594] Competition and Consumer Protection Commission, Digital Content / services. Available at: https://www.ccpc.ie/business/help-for-business/guidelines-for-business/digital-content-services/

[595] Ben-Hassine, Government Policy for the Internet Must Be Rights-Based and User-Centred. Available at: https://www.un.org/en/chronicle/article/government-policy-internet-must-be-rights-based-and-user-centred

[596] Department for Digital, Culture, Media & Sport, Consultation outcome: The government report on transparency reporting in relation to online harms (2020). Available at: https://www.gov.uk/government/consultations/online-harms-white-paper/outcome/government-transparency-report#key-findings-and-recommendations-by-theme

[597] Irish Council for Civil Liberties 2021: 'Free expression concerns with Online Safety and Media Regulation Bill.' Available at: https://www.iccl.ie/news/free-expression-concerns-with-online-safety-and-media-regulation-bill/

[598] Big Brother Watch, Index on Censorship, Open Rights Group, Article 19, Liberty, Adam Smith Institute, Re: Concerning developments for human rights online in the UK (2022). Available at: https://www.article19.org/wp-content/uploads/2022/11/Save-Online-Speech-Coalition-Letter-to-Special-Rapporteur-Final.pdf

[599] Ofcom, The Buffalo Attack: Implications for Online Safety (2022). Available at: https://www.ofcom.org.uk/__data/assets/pdf_file/0019/245305/The-Buffalo-Attack-Implications-for-Online-Safety.pdf

[600] Twitter, Permanent suspension of @realDonaldTrump (2021). Available at: https://blog.twitter.com/en_us/topics/company/2020/suspension

[601] Deutsche Welle, Merkel calls Trump Twitter ban 'problematic' (2021). Available at: https://www.dw.com/en/angela-merkel-calls-trump-twitter-ban-problematic/a-56197684

[602] Google, What is encryption? Available at: https://cloud.google.com/learn/what-is-encryption

[603] Wired, Apple Kills Its Plan to Scan Your Photos for CSAM. Here's What's Next (2021). Available at: https://www.wired.com/story/apple-photo-scanning-csam-communication-safety-messages/

[604] Apple, A Message to Our Customers (2016). Available at: https://www.apple.com/customer-letter/

[605] The Guardian, WhatsApp would not remove end-to-end encryption for UK law, says chief (2023). Available at: https://www.theguardian.com/technology/2023/mar/09/whatsapp-end-to-end-encryption-online-safety-bill

[606] BBC, Signal would 'walk' from UK if Online Safety Bill undermined encryption (2023). Available at: https://www.bbc.co.uk/news/technology-64584001

[607] Council of Europe, Human Rights for Internet Users.

[608] BBC, AI: War crimes evidence erased by social media platforms (2023). Available at: https://www.bbc.co.uk/news/technology-65755517

[609] Ben-Hassine, Government Policy for the Internet Must Be Rights-Based and User-Centred.

[610] Department for Digital, Culture, Media & Sport, Consultation outcome: The government report on transparency reporting in relation to online harms (2020).

[611] Mainly taken from: Department for Digital, Culture, Media & Sport, Consultation outcome: The government report on transparency reporting in relation to online harms (2020).

[612] European Union, Treaty on the Functioning of the European Union (2017). Available at: https://eur-lex.europa.eu/EN/legal-content/summary/treaty-on-the-functioning-of-the-european-union.html

[613] European Commission, The Digital Markets Act: ensuring fair and open digital markets. Available at: https://commission.europa.eu/strategy-and-policy/priorities-2019-2024/europe-fit-digital-age/digital-markets-act-ensuring-fair-and-open-digital-markets_en

[614] Thomson Reuters Practical Law, Competition Law Overview (Ireland) (2021). Available at: https://uk.practicallaw.thomsonreuters.com/w-028-2056?transitionType=Default&contextData=(sc.Default)&firstPage=true

[615] Charter of Fundamental Rights of the European Union (2000).

[616] CitizensInformation.ie, Charter of Fundamental Rights (2023).

[617] Charter of Fundamental Rights of the European Union (2000).

[618] Twitter, Permanent suspension of @realDonaldTrump (2021). Available at: https://blog.twitter.com/en_us/topics/company/2020/suspension

[619] European Commission, Mergers: Commission approves acquisition of WhatsApp by Facebook (2014). Available at: https://ec.europa.eu/commission/presscorner/detail/en/IP_14_1088

[620] European Commission, Antitrust: Commission sends Statement of Objections to Meta over abusive practices benefiting Facebook Marketplace (2022). Available at: https://ec.europa.eu/commission/presscorner/detail/en/ip_22_7728

[621] The Guardian, WhatsApp would not remove end-to-end encryption for UK law, says chief (2023). Available at: https://www.theguardian.com/technology/2023/mar/09/whatsapp-end-to-end-encryption-online-safety-bill

[622] BBC, Signal would ‘walk’ from UK if Online Safety Bill undermined encryption (2023). Available at: https://www.bbc.co.uk/news/technology-64584001

[623] Mainly taken from: Department for Digital, Culture, Media & Sport, Consultation outcome: The government report on transparency reporting in relation to online harms (2020).

Annex A | Further information on user rights

[624] The contents of this column are direct quotes from the Charter of Fundamental rights of the European Union. Charter of Fundamental Rights of the European Union (2000).

[625] The content in this column contains partial quotes from the Human Rights for Internet Users guide by the Council for Europe. Council of Europe, Human Rights for Internet Users. Available at: https://www.coe.int/en/web/freedom-expression/guide-to-human-rights-for-internet-users

[626] European Commission, Questions and Answers: Digital Services Act* (2023).

About PA

We believe in the power of ingenuity to build a positive human future.

As strategies, technologies and innovation collide we create opportunity from complexity.

Our diverse teams of experts combine innovative thinking and breakthrough use of technologies to progress further, faster, together. Our clients adapt and transform, and together we achieve enduring results.

We are over 4,000 strategists, innovators, designers, consultants, digital experts, scientists, engineers and technologists. And we have deep expertise in consumer and manufacturing, defence and security, energy and utilities, financial services, government and public services, health and life sciences, and transport.

Our teams operate globally from offices across the UK, Ireland, US, Nordics and Netherlands.

Discover more at paconsulting.com and connect with PA on LinkedIn and Twitter.

PA. Bringing Ingenuity to Life.

Corporate Headquarters

PA Consulting
10 Bressenden Place
London SW1E 5DN
United Kingdom

+44 20 7730 9000

paconsulting.com

This document has been prepared by PA. The contents of this document do not constitute any form of commitment or recommendation on the part of PA at the date of their preparation.

Produced by Design Studio at PA
(NB) 7_26007 December 2022.