Your response

Question 1: To assist us in categorising responses, please provide a description of your organisation, service or interest in protection of children online.
Is this a confidential response? (select as appropriate) No
5Rights develops new policy, creates innovative frameworks, develops technical standards, publishes research, challenges received narratives and ensure that children's rights and needs are recognised and prioritised in the digital world. While 5Rights works exclusively on behalf of and with children and young people under 18, our solutions and strategies are relevant to many other communities. Our focus is on implementable change and our work is cited and used widely around the world. We work with governments, inter-governmental institutions, professional associations, academics, businesses, and children, so that digital products and services can impact positively on the lived experiences of young people.

Question 2: Can you identify factors which might indicate that a service is likely to attract child users?
Is this a confidential response? (select as appropriate) No
Children must be protected wherever they are in practice, not only where government, companies and parents might wish them to be. One in three internet users worldwide are children.[1]The digital world is not optional for them. It is where they access education, health services and entertainment, build and maintain their relationships, and engage in civic and social activities. From search engines to social media sites to virtual reality spaces, children do not only use services explicitly targeted or designed for them. In order to protect children's rights, safety and to ensure their wellbeing, regulation must be geared towards services that children access in reality. The Online Safety Bill's scope is limited to user-to-user and search services. Regulated services that must comply with the Bill's child safety duties must be "likely to be accessed by children". This is defined as whether it is both possible for a child to access a service and whether the 'child user condition is met.'[2] This latter term is conditional on a

[1] Growing up in a connected world, UNICEF, link

[2] Clause 30 Children's Access Assessments, Online Safety Bill, link

Question 2: Can you identify factors which might indicate that a service is likely to attract child users?

"significant number of children" relative to the overall size of the user base of the service, or whether it is "likely to attract a significant number of users who are children."[3]

Ofcom must establish that services likely to attract children are not exclusively services designed for or targeted at children, in line with existing UK regulation. The thresholds established by the Data Protection Act (DPA)[4] and the Information Commissioner's Office Age Appropriate Design Code[5] should be mirrored in Ofcom's guidance to establish regulatory alignment across the Online Safety Bill with existing children's data protection regimes. This will ensure children are safe wherever they are online.

Section 123 of the DPA established a duty for the Information Commissioner to produce standards for age appropriate design of services. The Age Appropriate Design Code sets out how children's data should be managed as part of services data protection requirements, and specifically introduced a threshold condition of "likely to be accessed by a child."[6] The corresponding code makes clear that this threshold is not limited to services which are aimed at or are targeted at children. For a service to qualify as "likely to be accessed by a child", the possibility of a child accessing the service would need to be more probable than not. The Code further stipulates that certain features and characteristics may increase the chances of a child accessing a service.

Assessing whether or not a service is likely to be accessed by children allows providers to ensure children are given age-appropriate information, protections and experiences. The likelihood of children accessing a service can be determined by factors such as[7]:

- Whether its terms and conditions permit children to use the service
- Whether the service is directed at or predominantly used by children
- If children are already known to use the service, for example because a service already processes age-related personal data
- If children are known to use similar services
- The nature and content of the service and whether that has particular appeal for children
- The way in which the service is accessed and any measures that are (or are not) put in place to prevent children gaining access

If a service is unsuitable for children or children below a certain age, providers should focus on mechanisms which can prevent access. However, this should not lead to children being blocked from services they have a right to access, or to a service or certain features being downgraded for children.

[3] Clause 30 Children's Access Assessments, Online Safety Bill, link
[4] Data Protection Act 2018, link
[5] Age Appropriate Design Code 2020, link
[6] Age Appropriate Design Code 2020, link
[7] Age Appropriate Design Code 2020, link

Question 3: What information do services have about the age of users on different platforms (including children)?

Is this a confidential response? (select as appropriate)

No

Most services collect age data e.g., when a child registers for a social media account they are usually asked to enter their date of birth, their exact age, or to confirm they are above or below a certain age. Many also 'profile' users for a better understanding of their age, for example, by analysing the ages of 'friends' in the network, the type of content they engage with, their location, and the times of day they use the service.[8]

Many services ask users to consent to their data being shared with third parties. These third parties very often go on to share user data with their own partners, who may go on to share with their partners. As such, users signing up to a single service may have their age data shared with an endless chain of third parties. Some services allow users to register or sign in via existing accounts they have with other providers, such as Google, Apple, Facebook or TikTok.[9] Some providers make it difficult or even impossible to register or log in via other means. It is often unclear what information is shared between the service and the third party log in/authentication provider.

Question 4: How can services ensure that children cannot access a service, or a part of it?

Is this a confidential response? (select as appropriate)

No

Children have a right to access and participate in the digital world, and a right to access information which is not harmful to them and within the law[10][11]. Children must not be locked out of spaces they have a right to be in. Services must ensure they design features and systems with children's safety in mind. If a service is found to pose risk to children, it may be necessary to restrict a child's access to a service or a part of a service.

Children should not be able access to commercial pornography services due to the high risk this material poses to children. A service which hosts age-appropriate content to children but does host some mature content should not lock children out of their service but must be able to prevent children from encountering only that content and activity that may be harmful to them.

Age assurance is a mechanism with which services can ensure children are served age-appropriate experiences.[12] Age assurance describes any system or feature which pur-

[8] But how do they know it's a child?, 5Rights Foundation, link
[9] See Table A, 5Rights internal research
[10] UN Convention on the Rights of the Child, link
[11] UN General comment No. 25 (2021) on children's rights in relation to the digital environment, link
[12] But how do they know it's a child?, 5Rights Foundation, link

Question 4: How can services ensure that children cannot access a service, or a part of it?

ports to estimate or verify the age or age range of a user[13]. It is an umbrella term that captures a huge variety of approaches to ascertain age and encompasses both age estimation and age verification systems[14]. Age verification systems determine a person's age with a high level of certainty by checking against trusted, verifiable data[15]. Age estimation systems on the other hand estimate a person's age, using a combination of user provided data and algorithmic computation based on a large dataset[16]. Outputs vary from a binary determination as to whether someone is or is not above or below a certain age, through to placing an individual in a specific age category, through to estimating an exact age.

Different age assurance systems provide a sliding scale of confidence in the user's age. Systems can be used on their own or in combination to ensure higher confidence in the result. Combining age estimation systems can lead to very high levels of confidence in the final result[17]. Some common reasons for using age assurance are likely to be:

- To prevent underage users purchasing age-restricted goods
- To prevent underage users accessing or procuring age-restricted services
- To prevent underage users from viewing, accessing or consuming age-restricted content
- To provide age-appropriate experiences for different age groups

The level of assurance should be calibrated to the nature and level of risk presented by a product or service in relation to the age of the child. Crucially, age assurance must not be used to prevent children from participating in the digital world or to downgrade their experience. If a product or service meets all child safety requirements under the Online Safety Bill and is compliant with relevant data protection regulations, and is therefore appropriate for children of any age, there may be no need for age assurance. In general, less risky services will require a lower level of assurance. Services presenting a high risk to children, where the likelihood of harm to children occurring is high, or the impact of the harm is not minimal, including services required to comply with legal age limits, will need the highest bar of assurance.

Common approaches to age assurance include[18]:

- Self-declaration — requires a user to enter their birthdate or tick a box that asks if they meet the minimum age of use
- Hard identifiers — requires users to provide verified sources of identification to prove their age, such as a passport
- Biometrics — uses biometric information such as height, gait, voice, facial features, keystroke dynamics or finger and palm prints to identify a particular person or estimate their age
- Profiling and inference models — uses data from user behaviour to infer the age

[13] Ibid
[14] Ibid
[15] Ibid
[16] Ibid
[17] Ibid
[18] Ibid

Question 4: How can services ensure that children cannot access a service, or a part of it?

of users
- Capacity testing — estimates a user's age based on an assessment of their aptitude or capacity
- Account holder confirmation — requires a parent to confirm the age of a child user
- Device/operating system controls — offering controls on devices or through operating systems to deliver more age-appropriate experiences for children
- Flagging — allows users to 'flag' other users they believe do not meet a service's age requirements

Depending on the purpose, context and level of risk, services may implement a combination of approaches to age assurance.[19]

The Online Safety Bill cites age assurance as a method of how services may comply with their child safety duties[20]. However, there are currently no expectations of rules of the road for age assurance set out on the face of the Bill, leaving services without the clarity or direction they need to meet the child safety duties effectively. A lack of minimum standards may lead to the exacerbation of known problems of excessive data collection,[21] privacy infringements[22], ineffective age checks[23] and could lead to heavy-handed age-gating that can block children out of spaces they have a right to be in. It is key that age assurance is used in a way that is proportionate to the level of risk on their services and abides by the below principles:

- Age assurance must be privacy-preserving;
- Age assurance should be proportionate to risk and purpose;
- Age assurance should be easy for children to use;
- Age assurance not unduly restrict access of children to services to which they should reasonably have access, for example, news, health and education services;
- Age assurance providers must offer a high level of security;
- Age assurance providers must offer routes to challenge and redress;
- Age assurance must be accessible and inclusive;
- Age assurance must be transparent and accountable;
- Age assurance must be rights-respecting.

Any system of age assurance should not gather any more data than necessary about an individual to establish their age. Once that age or age range is established, the data used in the process of assurance should be stored or discarded transparently and securely.

The market and technology for age assurance systems is fast developing as demand for

[19] Ibid
[20] Clause 11 and Clause 25, Online Safety Bill, link
[21] Man files complaint accusing YouTube of harvesting UK children's data, The Guardian, link
[22] Largest FTC COPPA settlement requires Musical.ly to change its tune, Federal Trade Commission, link
[23] 60% of UK children aged 8-12 have a profile on at least one social media service, despite most social media having a minimum age requirement of 13, OFCOM, link

Question 4: How can services ensure that children cannot access a service, or a part of it?

these tools grows. Whatever the technology, age assurance systems can be privacy preserving if operated in accordance with standards of data minimisation and purpose limitation. Identity verification and age assurance should not be conflated. There is no need to confirm a user's identity when assuring their age. Children should not be routinely asked to disclose more information than is necessary to prove their age.

While age assurance is a useful tool for serving children age-appropriate experiences and preventing them from encountering harmful content and activity, age assurance alone is not sufficient for making a service age-appropriate for children. Action should be taken to mitigate risk, taking into account the ages of users and the particular risks posed by the service. For example, to make a service age appropriate a service some providers might simply need to disable some of their more intrusive or risky design features such as private messaging or livestreaming.

Question 5: What age assurance and age verification or related technologies are currently available to platforms to protect children from harmful content, and what is the impact and cost of using them?

Is this a confidential response? (select as appropriate)

No

The best approach to age assurance will be dependent upon the nature of the service being provided, the users that access the service, the type of content and activity on the service and the way that policies and terms and conditions are set out.

Age assurance mechanisms that are currently available include:

Self-declaration is often referred to as 'tick box' age assurance and is associated with the current failure to truly establish the age of children online. It requires a user to enter their birthdate, or to tick a box that asks if they meet the minimum age of use. However, when used in conjunction with additional proactive checks and technical measures, it can be much more effective. For example, where a child has submitted a date of birth that indicates they are above the minimum age, their provided age is checked again later in the process, such as when they next log in ("Can you remind us of your date of birth?"). Children who gave a false date of birth on registration may not remember the date of birth they gave when asked at a later stage or on a different day. Any discrepancy can be escalated to a moderator, who may ask for further proof of age.[24]

Hard identifiers for age assurance may require users to provide verified sources of identification to prove their age which contain many more attributes than age, such as name and address, or sensitive personal data such as race and gender. The more personal data that is captured to establish the age of users, the higher the standards of security, data retention and storage and accountability need to be. The use of hard identifiers offers

[24] But how do they know it's a child?, 5Rights Foundation, link

Question 5: What age assurance and age verification or related technologies are currently available to platforms to protect children from harmful content, and what is the impact and cost of using them?

a high level of assurance but presents risks of privacy violations and potential exclusion.[25]

Biometric data such as height, gait, voice, facial features, keystroke dynamics or finger and palm prints can be used to identify a particular person or to estimate their age through techniques such facial scanning, natural language processing and behavioural analysis.[26]

Profiling and inference models refers to the processing of data to analyse and infer information about a user, or to predict and determine aspects of user behaviour. Profiling for the purposes of age assurance is widely referred to as using 'AI' or 'inference models' to estimate age. Data used for profiling is made up from information users *choose* to share about themselves, and information that is *inferred* or automatically collected from their engagement with services, for example, how long they spend on a webpage, where their cursor hovers, the times of day they access a service and their interests, location and friends.[27]

Capacity testing allows a service to estimate a user's age based on an assessment of their aptitude or capacity. For example, a child may be asked to complete a language test, solve a puzzle or undertake a task that gives an indication of their age or age range. Services could use capacity testing to assure age without collecting personal data from children.[28]

Cross-account authentication is where a child uses an existing account to gain access to a new product or service. These accounts are often with large companies such as Apple, Facebook, Google or Twitter. In this form of authentication, the provider (the company) confirms that the user is who they say they are by asking them to enter the correct username and password for their account. Cross-account authentication can provide convenience for children by removing the need to prove their age every time they access a service, but as currently deployed it provides an unknown level of assurance. Without transparency around what data is shared, it also risks violating children's privacy and further embedding the market dominance of a handful of companies.[29]

Third party age assurance provider (ID/AV/AE). Third party age assurance providers are companies that offer age assurance or identity confirmation services. They work in multiple ways and offer services direct to users, such as digital IDs, or direct to businesses via API solutions, background checks, or tokenised age checking.[30]

Account holder confirmation enables a child's age or age range to be confirmed by an adult, often a parent or carer. Many forms of identification are only available to adults,

[25] Ibid
[26] Ibid
[27] Ibid
[28] Ibid
[29] Ibid
[30] Ibid

Question 5: What age assurance and age verification or related technologies are currently available to platforms to protect children from harmful content, and what is the impact and cost of using them?

such as a credit card or proof of eligibility to vote. This means that in many scenarios, it is easier for an adult to prove their age than it is for a child. Account holder confirmation leverages the knowledge that a service is likely to have about an adult user and gives the adult responsibility for confirming the age of the child.[31]

Device/operating system controls setups offer controls that may be designed to deliver more age appropriate experiences for children. They are generally limited to controls that restrict the websites and apps a child can access, filter content or set time-outs.[32]

Flagging is generally used to identify or 'flag' that something may be wrong. In the context of age assurance, it allows users to 'flag' other users they believe do not meet a service's age requirements. For example, a user of a dating app or subscription-based service selling adult content may be able to spot a user under the age of 18. Once a user is flagged to the service, their account can be blocked or a moderator may ask for proof of age.[33]

Question 6: Can you provide any evidence relating to the presence of content that is harmful to children on user-to-user and search services?

Is this a confidential response? (select as appropriate)

No

In 2021 5Rights commissioned research which explored how the design of digital products, in particular social media, shape the experiences and behaviours of children. Digital designers were interviewed to learn more about the design process and the commercial objectives that steer those processes. Children were interviewed about their experiences in the digital world and avatars based on real children's social media usage were created to replicate their experiences online. The findings formed the *Pathways: how digital design puts children at risk*[34] report which found:

Children were directly messaged and directed to pornographic content

- All of the child avatars were directly messaged by accounts on Instagram they did not follow. This included being added to group chats by strangers with other adults. The apparent motives behind these messages varied, but they included promoting websites with paid-for porn content, promoting brands or pages as well as offers to 'collaborate' in promoting products.
- All four male child avatars and two female child avatars on Instagram were added

[31] Ibid
[32] Ibid
[33] Ibid
[34] Pathways: How digital design put children at risk, 5Rights Foundation, link

Question 6: Can you provide any evidence relating to the presence of content that is harmful to children on user-to-user and search services?

to group chats by people they didn't know, in which there were multiple other strangers with links to paid-for porn sites or pornographic dating sites.

Children were nudged towards harmful and more extreme content:

- Ciara's (age 15) avatar searched for '#skinny' on Instagram, which led to an account titled 'skinny.quick' promoting a website selling 'fat burner' and 'breast enhancer' gummy bears.
- Charlotte's (age 15) avatar was recommended accounts and posts relating to weight loss and fitness. After 'following' and 'liking' a selection of these, the 'explore' feed filled with more similar content promoting weight loss journeys, fitness 'before and after' comparisons, dieting tips or photos of women emphasizing their slimness or low weight.

Children were easily able to search for and access harmful content

- Child avatars searched terms that aligned with content that children in the research had told us they had seen on social media apps and sites, including 'thin', 'bodygoals', 'porn', 'darkmemes', 'suicide' and 'proanaa' (proana with one 'a' is blocked by Instagram, but adding a second 'a' unlocks access to pro-anorexia content).
- Laura's (age 13) avatar searching 'suicide' on Instagram was recommended images of graphic self-harm injuries
- Ciara's (age 15) avatar searching for 'proanna' on Instagram was recommended posts titled "no food all week"
- Jordan's (age 14) avatar was served up sexual content on Instagram alongside adverts for Roblox and a school revision study app
- Owen's (age 15) avatar was served sexual content on Instagram alongside adverts for T-levels and a Home Office campaign for recognising and reporting child abuse online
- Laura's (age 13) avatar was able to search for 'depressed' theme content on Instagram, while also being served adverts for a sweetshop, Nintendo Switch and a teen targeted-tampon advert. Laura's avatar was also recommended a post of pro-suicide material saying "it's so easy to end it all."

Children also encounter commercial harms from in-game ads and persuasive design features. Loot boxes present a serious risk of financial harm to young people, which can have devastating consequences for both young people and their families.[35]

British children collectively spent £270 million pounds on mobile game loot boxes and other in-app purchases in 2019.[36] But young people feel that the lack of safeguards against financial harms are intentional, with 76% of young video game players believing that online video games try to make you spend as much money as possible.[37]

[35] Loot boxes: I blew my university savings gaming on Fifa, BBC News, link
[36] Resisting the Irresistible: In-app purchases & Loot Boxes, TechSafe, link
[37] The Rip-Off Games: How the new business model of online gaming exploits children, Parent Zone, link

Question 6: Can you provide any evidence relating to the presence of content that is harmful to children on user-to-user and search services?

A BBC report documented cases of young people spending up to £2,000 on in-app gaming purchases,[38] where there has been insufficient friction or safeguarding against spending. Such cases of young people being able to complete multiple transactions within the same gaming session, or re-purchase and download the same items over-and-over again, indicate that current measures against financial safeguarding are inadequate.

Question 7: Can you provide any evidence relating to the impact on children from accessing content that is harmful to them?

Is this a confidential response? (select as appropriate)

No

It is important to recognise that not all children are the same, and children's experiences online are shaped by multiple environmental and personal factors, including age, developmental capacity, gender, sexuality, and familial circumstances. It follows that there will be certain groups of children, at certain times, who will be more vulnerable to the effects of harmful content. Children will experience periods of increased sensitivity relative to their developmental stage, but not necessarily at the same age. In girls, for example, social media use between the ages of 11 and 13 years is associated with a decrease in life satisfaction one year later, whereas in boys this happens later between the ages of 14 and 15 years, suggesting that sensitivity to social media is linked to developmental changes that occur later in boys than girls.[39]

Whilst services emphasize correlation should not be conflated with causation in relation to rising social media use and lower levels of life satisfaction and wellbeing among teens, it has not escaped the attention of researchers, parents and children themselves that the arrival and widespread adoption of social media between 2009 and 2012 coincided with a "collapse in teen mental health[40]" that continues today.[41] From 2009 to 2019, the proportion of US high school students reporting persistent feelings of sadness or hopelessness increased by 40%.[42] Mounting evidence of the positive correlation between high social media use and mood disorders has refocused attention on those aspects of the

[38] My son spent £3,160 in one game, BBC News, link

[39] Windows of developmental sensitivity to social media, Orben, A., Przybylski, A.K., Blakemore, SJ. et al Nature Communications 13, 1649, link

[40] Testimony of Professor Jonathan Haidt before the Senate Judiciary Committee, Subcommittee on Technology, Privacy, and the Law May 4, 2022, Senate Judiciary, link

[41] The number of US teenagers with depression doubled between 2011 and 2019, with more than *23%* of girls ages 12-17 experienced a major depressive episode during 2019, How Much Is Social Media to Blame for Teens' Declining Mental Health?, Institute for Family Studies, link

[42] Protecting Youth Mental Health, The U.S surgeon General's Advisory, link

Question 7: Can you provide any evidence relating to the impact on children from accessing content that is harmful to them?

online world, including persuasive design, that may contribute to lower wellbeing, and exacerbate existing mental disorders among teens.[43] Managing public and frequent interactions online creates enormous pressures for young people, and with it comes anxiety, low self-esteem and mental health challenges at ever-increasing levels.[44] Research from Internet Matters shows a clear pattern between spending more time online and being more likely to experience an online harm. Their research reveals that "children who say they have been highly affected by their experience of an online harm were more than three times more likely to be using chatrooms and forums (+243% difference) and creators (uploading videos they made themselves, +218%) compared to those children who haven't experienced any online harm."[45]

The impacts of harmful content include but are not limited to:

- Immediate or cumulative
- Acute or mild
- Direct and Indirect

Immediate and Cumulative

The impact of content can be immediate if the content is graphic, whereas content that is non-egregious in nature but can be harmful in large volumes may have a cumulative impact on children. A steady drip feed of body-image focused content accompanies by virality, and high engagement such as likes and comments can normalize unrealistic body types or sexualized images of young women and girls. According to Meta's own internal research, a third of teenage girls also believed that Instagram made them feel worse about those their bodies.[46]

Molly Russell was 14 when she ended her life after viewing graphic self-harm, suicide and depression related content on social media. The coroner leading the inquest into Molly's death concluded that she had died from an act of self-harm while suffering from depression and "the negative effects of online content."[47] The coroner observed that while the content itself was harmful, it was made considerably worse by features such as comments, hashtags and likes, with some posts attracting over 10,000 likes.[48] High numbers of likes and comments created a sense of legitimacy and normalised the extreme content, and as the coroner noted "glamorised and even glorified" self-harm.[49] This illustrates how a large volume of pro-self-harm and pro-suicide content directed at or accessed by children can have a significant cumulative impact.

Acute and mild

Content may have an acute or mild impact on a child according to the child and their cir-

[43] Problematic smartphone use: a conceptual overview and systematic review of relations with anxiety and depression psychopathology, J Elhai et al, Journal of Affective Disorders 207, 251-259, link
[44] Mental Health of Children and Young People in England 2022 - wave 3 follow up to the 2017, NHS, link
[45] Exploring the impact of online harms, Part 2, Internet Matters, link
[46] Facebook Knows Instagram Is Toxic for Teen Girls, Company Documents Show, Wall Street Journal, link
[47] REGULATION 28 REPORT TO PREVENT FUTURE DEATHS, Senior coroner Andrew Walker, link
[48] Molly Russell Inquest proceedings, 2022
[49] Molly Russell Inquest proceedings, 2022

Question 7: Can you provide any evidence relating to the impact on children from accessing content that is harmful to them?

cumstances and characteristics. Factors such as gender must be taken into consideration when assessing the impact on content that is harmful to children. Many studies have found that the correlation between social media use and harm is stronger among girls. Meta's own internal research, revealed by whistle-blower Frances Haugen, showed that among teenage girls experiencing suicidal thoughts, 6% in the US and 13% in the UK traced those thoughts back to Instagram.[50]

As the difference between 'online' and 'offline' has become less distinct, the negative effects of children's offline experiences carry over into their 'online' lives and vice versa. Content and interactions online that feel more personal, familiar or local can impact a young person's 'offline' behaviour. Ofcom's own report examining online risk factors included testimony from one 13-year-old girl who now avoids people and places in her local area due to content she sees online.[51]

An excessive amount of sharing can also lead to exaggerated, polarising and aggressive behaviour, fuelled by the need to be noticed. Research from the Youth Endowment Fund in 2022 found that 55% of teens had seen real life acts of violence on social media in the last 12 months, and 62% who had committed an act of violence themselves thought social media played a major role in why children commit violence.[52]

Direct and indirect

The way in which a child is interacting with the content can influence the content's impact on them. If a child is directly viewing or experiencing the content or the content is about them the impacts will likely be more substantive.

1 in 4 girls in the UK have shared a sexual image of themselves and of those, a quarter said they felt pressured into it, and almost a third initially wanted to but later regretted it.[53] The spreading of Child Sexual Abuse Material (CSAM), revenge porn and nude imagery can often retraumatize the victims.

The digital environment in which the content is being experienced or generated in can also shape how direct or indirectly the impacts of harm are felt. While many of the risks to children in the metaverse are the same as those found in other digital spaces, such as social media or games, there are certain risks that are unique to or exacerbated in the metaverse. Heightened sensory experience through the use of haptic technologies (beyond text and image interaction) creates a different sensorial experience with the potential to intensify feelings of emotional and physical distress. If a child's avatar is physically assaulted or if a stranger whispers into their avatar's ear this will have a more direct impact on the child as the interaction is immersive and personal. Psychotherapist Nina Jane Patel logged into Horizon Venues metaverse which was launched by Facebook in 2020 and created her avatar. "Within 60 seconds of joining — I was verbally and sexually

[50] Facebook Knows Instagram Is Toxic for Teen Girls, Company Documents Show, Wall Street Journal, link

[51] Research into risk factors that may lead children to harm online, Ofcom and Revealing Reality, link

[52] Children, violence and vulnerability 2022, Youth Endowment Fund, link

[53] Sexual harassment at school: new film co-created with young people supported by latest data, End Violence Against Women, link

Question 7: Can you provide any evidence relating to the impact on children from accessing content that is harmful to them?

harassed — 3–4 male avatars, with male voices, essentially, but virtually gang-raped my avatar and took photos — as I tried to get away they yelled — 'don't pretend you didn't love it'."[54] She has since documented having panic attacks and mental health consequences as a result of this assault.

The use of avatars will exacerbate problems associated with the 'filter' effect of social media beautifying features, creating unrealistic body ideals and pressure among children to look a certain way. Mitch Prinstein, a clinical psychologist who serves as chief science officer for the American Psychological Association said of the Metaverse: "this is just an exacerbation of the problems that we've already started to see with the effects of social media…This is creating more loneliness. This is creating far more body image concerns [and] exposure to dangerous content that's related to suicidality." He noted that being able to modify your likeness to project a version of yourself that differs from real life can be "pretty dangerous for adolescents, in particular."[55]

A survey by the NSPCC and the Children's Commissioner for England found that 44% of boys aged between 11 and 16 who regularly viewed pornographic content reported that it gave them ideas about the type of sex that they wanted to try.[56] Most young people also said girls expect sex to involve physical aggression, such as airway restriction.[57] This corroborates findings from the UK school's regulator as part of its review into sexual abuse in schools, which found that "children and young people… had learned more about sexuality from social media than from school or had got their education about relationships from their peers and social media."[58]

Ofcom's own research shows how these categories may intersect. Cumulative active engagement with hazards over time can self-reinforce behaviour to cause significant and severe harm such as engaging with and participating in pro-anorexia communities online. Cumulative passive exposure to hazards over time that can build up to cause more significant harm such as being immersed in body-focused content in social media feeds.[59]

Services design features also impact how children experience the content on-site. A feature aimed at creating false scarcity to increase immediate user engagement will be felt by the child as pressure to be online. 5Right's 2021 *Pathway's* research found children were concerned about their online experiences which mirrored the design strategies behind the services' features.[60] These included:

- Spending too much time online
- Being contacted by adult strangers
- Feeling pressure to behave in ways that gain attention and
- validation
- Feeling pressure to be 'visible' and active online

[54] Reality or Fiction? 21 Dec 2021, Nina Jane Patel, link
[55] This is creating more loneliness': The metaverse could be a serious problem for kids, experts say, CNBC, link
[56] 'A lot of it is actually just abuse'- Young people and pornography, Children's Commissioner for England, link
[57] Ibid
[58] Review of sexual abuse in schools and colleges, Ofsted, link
[59] Research into risk factors that may lead children to harm online, Ofcom, link
[60] Pathways: How digital design put children at risk,5Rights Foundation, link

Question 7: Can you provide any evidence relating to the impact on children from accessing content that is harmful to them?

- Enhancing their appearance through image alteration

The young people 5Rights works with have described their experience of harmful content and what they think service providers should change:

Haydn, 16: "One thing I'd change is making it harder to access inappropriate content in the sense of really violent videos and other stuff. Because it's just really easy to find that stuff."

Liam, 14: "I would make sure that hate comments were the priority of getting rid of. I see so many of those and it really does impact people's mental health quite a lot."

Niamh, 15: "I feel like for people who just want to post videos for fun, obviously maybe they should have a private account. But if they want to publicly post videos, especially younger people who just want to have fun on the app, I feel like it can be quite harmful if the comments are negative especially. It can be really dangerous and I think it's a good idea to limit how viral something can go."

Question 8: How do services currently assess the risk of harm to children in the UK from content that is harmful to them?

Is this a confidential response? (select as appropriate)

[Please select]

Question 9: What are the exacerbating risk factors services do or should consider which may have an impact on the risk of harm to children in the UK?

Is this a confidential response? (select as appropriate)

No

Question 9: What are the exacerbating risk factors services do or should consider which may have an impact on the risk of harm to children in the UK?

In considering how children can be exposed to the same material or activity and yet experience it differently, consider the following elements:

- Hazards: the content, contact, conduct or commercial relationship that the child encounters
- Risk factors that influence the likelihood of a child experiencing the hazard as harmful, including their personal characteristics, their circumstances and the design of the service
- Harms: the negative outcomes that occur after exposure to a hazard, and shaped by risk factors

It is also important to remember that not all children are the same, and do not always develop at the same rates. Some of the young people we work with recognise that children have different needs and vulnerabilities:

Vivien, 16: "We all have different emotional intelligence. All services should assume that, so everyone is kept safe rather than a few."

Manahil, 17: "Different age groups have different needs and wants. That could be assessed by referring back to their age, so it's important that different age groups have different features to meet their needs."

New research from indicates that social media use among girls between the ages of 11 and 13 years is associated with a decrease in life satisfaction the following year, whereas in boys this happens later between the ages of 14 and 15 years, suggesting that sensitivity to social media is linked to developmental changes that occur later in boys than girls.[61]

Using features in combination

Providers should recognise how the risks created by individual features can increase when they are used in combination with other features. For instance, a service which makes use of algorithmic friend recommendations that recommend child accounts to adults and make a child's location discoverable by other users would make the potential for grooming and sharing CSAM more likely.

Misuse of features

Providers should account for how their features might be misused by actors with malign intent. Such risks may arise through:

- Inauthentic use of the service, such as the creation of fake accounts
- The use of bots or deceptive use of the service
- Other automated or partially automated behaviours
- Coordinated manipulation and use of their services

[61] Windows of developmental sensitivity to social media. Orben, A., Przybylski, A.K., Blakemore, SJ. et al Nature Communications 13, 1649, link

Question 9: What are the exacerbating risk factors services do or should consider which may have an impact on the risk of harm to children in the UK?

- Systemic infringement of their terms of service

Providers should pay particular attention to how their services, and any use of algorithmic amplification, may contribute to these systemic risks.

Risks over time

Certain risks may expose children to low levels of immediate harm but increase in severity over time. A single notification may momentarily distract a child, for instance, but over time may have a more serious impact on their sleep, schoolwork and ability to concentrate.

Providers should consider both the immediate and longer-term impacts their features have on children.

Children can be exposed to risks over time in many ways, including:

1. Isolated exposure to risks that cause immediate harm, such as seeing a violent, sexual or otherwise developmentally inappropriate content
2. Cumulative passive exposure to risks over time, such as seeing the same narrow ideals of beauty consistently promoted in newsfeeds or timelines
3. Cumulative active engagement with risks, such as participating in pro-anorexia or self-harm groups

Similarly, the impact of harm can be either:

- Immediate or delayed – whether the impact of the experience occurred immediately after exposure or manifested at a later point
- Direct or indirect – whether the impact of the hazard occurred through direct exposure to the child who was harmed or indirectly through exposure

Children's vulnerability to harm

Service providers should consider the particular vulnerabilities of different children, at different development stages. For example, some children may be sharing devices with older siblings, and a six-year-old will require different protection measures than a 16-year-old. Other factors to consider include, but are not limited to:

- Lack of digital access
- Low self-esteem
- Cognitive development issues
- Mental or physical illness
- Having previously been a victim or perpetrator of conduct harms
- Lack of parental or guardian support
- Socio-economic deprivation
- Family difficulties
- Disability
- Educational disadvantage

Question 9: What are the exacerbating risk factors services do or should consider which may have an impact on the risk of harm to children in the UK?

Risks to groups and society

As well as presenting risks to individual children, products and services might also pose risks to certain groups and wider society. For instance, the abuse of women online may have a chilling effect on girls' self- expression.[62] Automated decision-making, has been shown to discriminate against certain groups when trained on poor datasets, such as those trained only on adults or on a particular gender or race.[63] Young people's trust in democratic processes and institutions is undermined if services that encourage virality also allow disinformation to spread widely. Providers have a responsibility to consider and address these collective and societal risks that impact children, as well as risks experienced by children as individuals.

Question 10: What are the governance, accountability and decision-making structures for child user and platform safety?

Is this a confidential response? (select as appropriate)

No

5Rights Foundation and the Institute of Electrical and Electronics Engineers Standards Association collaborated to develop the pioneering IEEE 2089-2021 Standard for Age Appropriate Digital Service Framework.[64] The Standard introduces practical steps that companies can follow to design digital products and services that are age appropriate.

Roles and responsibilities for ensuring a high level of privacy, safety, and security for children must be established within the organisation. The competencies suggested by the IEEE 2089-2001[65] include:

- Technical domain knowledge: empirical, academic or a blend of both
- Experience of application (knowing what works) in different contexts and the requisite skills
- The drive and motivation to achieve the goals and strive for improvement or excellence
- Having appropriate behaviours, such as teamwork, leadership, and compliance with professional codes
- The ability to adapt to changing circumstances and demands by creating new know-how

[62] No space for violence against women and girls in the digital world, Council of Europe, link
[63] AI bias: exploring discriminatory algorithmic decision-making models and the application of possible machine-centric solutions adapted from the pharmaceutical industry, Belenguer, L., *AI Ethics* **2**, 771–787 (2022), link
[64] IEEE Standard for an Age Appropriate Digital Services Framework Based on the 5Rights Principles for Children, IEEE Consumer Technology Society, link
[65] Ibid

Question 10: What are the governance, accountability and decision-making structures for child user and platform safety?

- The ability to perform the requisite tasks efficiently and reduce waste of physical and virtual resources
- The ability to understand the needs of stakeholders and deliver a high-quality service
- A commitment to creating an age-appropriate product or service

Whatever the size of an organisation, the functions of the following roles must be fulfilled by one or more people with the correct expertise and commitment to upholding children's rights, either within the organisation or from a specialist third-party provider. This should include a commitment to meeting the diverse needs of children and to tackling gender-based risks.

Roles which could be considered are:

Children's safety management champion

This role will support the establishment of a safety culture by prioritizing children's safety in organisational values, hiring processes and core business models.

The responsibilities for this role include establishing and upholding a corporate commitment to providing a safe service for children. This applies throughout the lifecycle of any product or service that is likely to be accessed by or impact children. For example, ensuring a team prioritises child-centred design and resolving any conflicts between this overall objective and corporate strategies. The children's safety management champion works closely with the child safety lead and child rights advocate, providing a point of contact and expertise.

They are responsible for responding to concerns from those involved in the development and deployment of age-appropriate products and services.

Child safety lead

This person is responsible for providing knowledge and understanding of the 4Cs of online risk[66], children's online safety and children's rights, to the teams responsible for product design and development. The child safety lead will facilitate discussions in related activities, for example providing specific training and development, working closely with the transparency manager to support the child impact assessment process. This role leads the vision to create service and product design, directing the implementation of decisions within engineering, user experience design, marketing and outreach, policy compliance and/or customer support teams.

Child rights advocate

This person ensures a child's rights perspective is taken to the development of products or services with the aim of embedding child-centred design and resolving conflicts in the best interests of children. As part of this, the child rights advocate will facilitate participa-

[66] The 4Cs: Classifying Online Risk to Children. (CO:RE Short Report Series on Key Topics). Hamburg: Leibniz-Institut für Medienforschung | Hans-Bredow-Institut (HBI); CO:RE - Children Online: Research and Evidence, Livingstone, S., & Stoilova, M. (2021), link

Question 10: What are the governance, accountability and decision-making structures for child user and platform safety?

tory design with children and represent children in project team meetings in cases when they cannot be directly involved.

The responsibilities for this role include listening to stakeholders and team members to understand concerns and consider a range of technical solutions. The child rights advocate will liaise with teams responsible for systems and process to ensure that safety is baked in by design and default, while maintaining a commitment to child-centred design and the best interests of the child. They are responsible for leading the child impact assessment process and for developing subsequent plans for risk mitigation, with responsibility for the integrity of the assessment process.

Transparency manager

The responsibilities for this role include producing regular external reports that provide detail on key aspects of the service. These may include specific insights into the intention, functionality and consequences of algorithms, the number of accounts created by underage users removed, and the measures a service has taken to protect child users. The transparency manager is also responsible for internal record keeping, ensuring that transparency considerations are made at regular intervals and crucial milestones. They need to ensure that decisions are recorded in a consistent form so they are easily retrievable and key information can be accessed, such as the names of those accountable for decisions. The transparency manager is also responsible for leading the child impact assessment process and for developing subsequent plans for risk mitigation, with responsibility for the integrity of the assessment process.

Question 11: What can providers of online services do to enhance the clarity and accessibility of terms of service and public policy statements for children (including children of different ages)?

Is this a confidential response? (select as appropriate)

No

Under the Online Safety Bill providers will be required to make their terms clear and accessible.[67] To make published terms age-appropriate, organisations must consider the following:[68]

Language

Published terms should avoid jargon and spell out key definitions and terms. The language used should be simple, straightforward and pitched at a level that the youngest likely user can understand, or presented in different versions to suit different age groups.

[67] Clause 11 and 25, the Online Safety Bill, link

[68] Tick to Agree, 5Rights Foundation, link

Question 11: What can providers of online services do to enhance the clarity and accessibility of terms of service and public policy statements for children (including children of different ages)?

Length
Published terms should be concise and to the point. They should be short in length (word count), divided into clear sections or made available in bite-sized pieces.

Format
Key terms and definitions should be prominent, presented in bold text or graphics and icons if needed. Consulting with children on the most appropriate format and testing these with diverse groups of children enables their views to be heard and can guide formatting and design decisions.

Navigability
Published terms should be prominent and easy to find and key terms and definitions should also be searchable.

Timing
Published terms should be presented at multiple or significant times in the user journey. Ongoing, meaningful engagement at regular intervals and at crucial moments, including every instance where consent is sought, can support a child to comprehend any terms of agreement they are entering into.

Accessibility
Published terms should consider the diverse needs of young people. This includes providing terms in multiple languages and catering for children with accessibility needs. Providers should not assume children have an engaged adult on hand to help them understand terms.

Ensuring meaningful consent
Consent must be sought and obtained, not assumed, and must be given by a "clear affirmative act establishing a freely given, specific, informed and unambiguous indication.[69]" Obtaining meaningful consent means that a child understands and accepts to terms at all stages of their user journey and may choose to change their mind at a later point. 'Tick box' or 'unread' consent must not be used when the end user is a child.[70]

Children must be given the option to refuse individual terms without being precluded access to other parts of the service. Where parental consent is required, this should be meaningful and steps should be taken to verify that the parent or guardian is who they say they are.

Providers should seek consent whenever amendments are made to the service, explaining the changes and their implications for the user, and it should be possible for users to withdraw consent, both after regular periods of time and at times of their own choosing.

Upholding published terms

[69] What is valid consent, ICO, link
[70] Tick to Agree, 5Rights Foundation, link

Question 11: What can providers of online services do to enhance the clarity and accessibility of terms of service and public policy statements for children (including children of different ages)?

Terms must be consistently enforced to create a culture of good governance and clarity for parents and young people about what constitutes a violation of service use agreements. Redress and reporting information must be prominent and easily accessible. It must also be clear what happens when a user makes a complaint. Expectations of response times must be clearly set out in the terms and upheld by the provider, and reports relating to young people's safety should take priority. Organisations should consult stage 12 of IEEE Standard 2089 for more details on making published terms age-appropriate.[71]

Question 12: How do terms of service or public policy statements treat 'primary priority' and 'priority' harmful content?

Is this a confidential response? (select as appropriate)

[Please select]

Question 13: What can providers of online services do to enhance children's accessibility and awareness of reporting and complaints mechanisms?

Is this a confidential response? (select as appropriate)

No

[71] IEEE Standard for an Age Appropriate Digital Services Framework Based on the 5Rights Principles for Children, IEEE Consumer Technology Society, link

Question 13: What can providers of online services do to enhance children's accessibility and awareness of reporting and complaints mechanisms?

Not all children are the same; they are different ages, have different capacities, speak different languages and exist in different familial and social circumstances. The following factors should be considered when making a product or service inclusive and accessible[72]:

- The needs of children with disabilities
- The age or age range of the child
- The needs of children who may not have active or engaged parents or guardians
- The needs of vulnerable groups and children with protected characteristics
- The affordability of the product or service

Services should be designed to meet the needs of children in a wide range of circumstances, and at a minimum, all the children who are likely to access them. This includes:

- Providing content in local languages, including information about moderation and redress
- Providing material that reflects the full range of cultures and conditions of the users
- Meeting accessibility requirements and standards, including the latest Web Content Accessibility Guidelines

Question 14: Can you provide any evidence or information about the best practices for accurate reporting and/or complaints mechanisms in place for legal content that is harmful to children, or users who post this content, and how these processes are designed and maintained?

Is this a confidential response? (select as appropriate)

[Please select]

[72] IEEE Standard for an Age Appropriate Digital Services Framework Based on the 5Rights Principles for Children, IEEE Consumer Technology Society, link

Question 15: What actions do or should services take in response to reports or complaints about online content harmful to children (including complaints from children)?

Is this a confidential response? (select as appropriate)

No

Providers should:[73]

- Provide prominent, accessible and easy-to-use tools to help children and parents seek redress, including by highlighting how to use them during the sign-up/ induction process and tailoring tools to the age of the child
- Provide children and parents access to expert advice to support their decision-making and help them understand their rights
- Have clear penalties applied fairly and consistently
- Offer opportunities to appeal decisions, and escalate unresolved appeals to expert third parties or regulators
- Provide response times that are appropriate to the seriousness of the report being made, including by responding immediately to children who appear to be in distress
- Provide children and parents with opportunities to correct a child's digital profile/footprint, with clear and accessible tools that match up to a child's data rights
- Inform children of action taken in redress processes by granting access to the status of their reports, communicating actions clearly and giving them the opportunity to provide feedback

Question 16: What functionalities or features currently exist that are designed to prevent or mitigate the risk or impact of content that is harmful to children? A1.21 in the call for evidence provides some examples of functionalities.

Is this a confidential response? (select as appropriate)

[Please select]

[73] IEEE Standard for an Age Appropriate Digital Services Framework Based on the 5Rights Principles for Children, IEEE Consumer Technology Society, link

Question 17: To what extent does or can a service adopt functionalities or features, designed to mitigate the risk or impact of content that is harmful to children on that service?

Is this a confidential response? (select as appropriate)

No

Providers should assess the risks outlined in the 4Cs framework[74] presented by each feature of the product or service to reveal known harms, potential risks and unintended consequences. Providers should consider the potential for their own features to negatively impact children, with attention paid to how features may be experienced differently when encountered in combination, and how they might impact different groups of children. At the end of this process, providers will be able to identify elements or features that may need to be disabled, redesigned or carry warnings and/or other mitigation measures in order to keep children safe. It will also allow providers to make positive changes that deliver enhanced, creative and age-appropriate experiences.

Providers should consider both the likelihood of harm occurring and the severity of harm when it does occur. The likelihood of a child encountering harm can be measured by, among other methods, peer-reviewed academic research, internal research, A/B testing and data from public bodies. Providers should make use of child development experts, official advice from public health authorities and the testimony of children themselves when measuring the severity of harm.

Providers should recognise how the risks created by individual features can increase when they are used in combination with other features. For instance, a service which makes use of algorithmic friend recommendations that recommend child accounts to adults and make a child's location discoverable by other users could make the potential for grooming and sharing Child Sexual Exploitation and Abuse (CSEA) more likely.

Question 18: How can services support the safety and wellbeing of UK child users as regards to content that is harmful to them?

Is this a confidential response? (select as appropriate)

No

[74] The 4Cs: Classifying Online Risk to Children. (CO:RE Short Report Series on Key Topics). Hamburg: Leibniz-Institut für Medienforschung | Hans-Bredow-Institut (HBI); CO:RE - Children Online: Research and Evidence, Livingstone, S., & Stoilova, M. (2021), link

Question 18: How can services support the safety and wellbeing of UK child users as regards to content that is harmful to them?

Assessing risk requires consideration of several interlocking and interdependent factors. Providers should consider the complexity of risk when assessing their service and features. Approaches to risk assessment and mitigation should be based on "the best available information and scientific insights." In particular, providers should think about:

- How certain features in combination may exacerbate risk
- How some risks are not immediately obvious but may create significant harm over time
- How children with different levels of vulnerability or resilience may respond
- How both individuals and groups experience risk

Service providers must assess for the full range of harms (content, contact, conduct and contract harms[75]) children experience online many of which are created by the design features and commercial decisions of platforms, which are often the true drivers of harm. For example, it is often the pursuit of maximising user-engagement that fuels the amplification and targeting of harmful content to children.[76]

Services must mitigate and manage the resulting risks and design their services with children's safety in mind. They must uphold children's existing rights to participate and express themselves in the digital world as per the United Nations Convention on the Rights of the Child[77] and General Comment No.25 on children's rights in relation to the digital environment.[78]

Question 19: With reference to content that is harmful to children, how can a service mitigate any risks to children posed by the design of algorithms that support the function of the service (e.g. search engines, or social and content recommender systems)?

Is this a confidential response? (select as appropriate)

No

Recommendation systems are fundamental to the design and operation of most online services, from social media to search engines. They are optimised to increase user engagement, and ultimately to fulfil the commercial objectives of the provider. 5Rights research[40] has shown that designers of such systems are tasked with achieving three main objectives:

- To maximise time spent on the platform

[75] The 4Cs: Classifying Online Risk to Children. (CO:RE Short Report Series on Key Topics). Hamburg: Leibniz-Institut für Medienforschung | Hans-Bredow-Institut (HBI); CO:RE - Children Online: Research and Evidence, Livingstone, S., & Stoilova, M. (2021), link

[76] Pathways: How digital design put children at risk, 5Rights Foundation, link

[77] UN Convention on the Rights of the Child, Office of the High Commissioner on Human Rights, link

[78] UN General comment No. 25 (2021) on children's rights in relation to the digital environment, Office of the High Commissioner on Human Rights link

Question 19: With reference to content that is harmful to children, how can a service mitigate any risks to children posed by the design of algorithms that support the function of the service (e.g. search engines, or social and content recommender systems)?

- To maximise reach by attracting as many users as possible
- To maximise activity, encouraging users to generate and interact with as much content as possible

The prioritisation of these objectives leads to recommendation systems serving up content that is most likely to extend user engagement, regardless of whether that content is potentially harmful to users.

Providers should follow best practice when designing and operating recommendation systems, as set out below. With robust risk mitigation measures in place, greater transparency and effective oversight, the risks posed to children by recommendation systems can be reduced. Services can mitigate risks to children if they, for example:

- Assess the impact of algorithms used in recommendation systems, considering the objectives, data inputs, the rules which weight information with more or less importance, and the intended and unintended outcomes[79]
- Don't target children with advertising and make clear when content is sponsored or paid-for
- Do not compare children with adults for 'people also liked…' features
- Give children a variety of clear and accessible options to prioritise the type of posts they want to see or to turn off 'personalisation' altogether
- Place less importance on 'popularity' and 'performance' when ranking recommendations, to broaden the variety and strengthen the veracity of information children are able to access
- Don't recommend offensive or age-inappropriate suggestions for autocomplete
- Provide information on how content has been ranked, showing the data and algorithms used to arrive at a decision
- Restrict adults from viewing children's accounts in friend or follower recommendations, and don't show young people's content to adults as suggested content
- Implement privacy-preserving and effective age assurance so children can be provided with age-appropriate experiences

Question 20: Could improvements be made to content moderation to deliver greater protection for children, without unduly restricting user activity? If so, what?

Is this a confidential response? (select as appropriate)

No

[79] IEEE Standard for an Age Appropriate Digital Services Framework Based on the 5Rights Principles for Children, IEEE Consumer Technology Society, link

Question 20: Could improvements be made to content moderation to deliver greater protection for children, without unduly restricting user activity? If so, what?

Features that enable users to block, limit or filter certain keywords or types of content can limit their exposure to harmful content. Users should be given the option of switching off comments and recommended content. For instance, Instagram has introduced two new features, *Favourites* and *Following*, to offer users more choice over what they see on their news feeds. *Favourites* allows users to see posts from accounts they have selected as those they are most interested in, while *Following* shows recent posts from all accounts a user follows. Neither feature has recommended content, and both are available as additional tabs on the homepage.

Many services also make use of just-in-time warnings, informing users of potential risks associated with content they are about to interact with.

Features which add friction to sharing may also help prevent the risks of virality, where potentially harmful content reaches large numbers of people before it can be moderated. For instance, allowing content to be forwarded to a limited number of people reduces the risk of harmful content being sent on to everyone in a user's contact list.

Services likely to appeal to younger age groups should consider pre-moderating content, screening all potential uploads before it can be viewed by other users.

5Rights often hears from young people that they would like to have age-appropriate empowerment measures alongside safety features, to enable them to manage their own experiences. Such measures will be especially relevant for older children, who have greater autonomy as they mature.

Question 21: What automated, or partially automated, moderation systems are currently available (or in development) for content that is harmful to children?

Is this a confidential response? (select as appropriate)

[Please select]

Question 22: How are human moderators used to identify and assess content that is harmful to children?

Question 22: How are human moderators used to identify and assess content that is harmful to children?

Is this a confidential response? (select as appropriate)

[Please select]

Question 23: What training and support is or should be provided to moderators?

Is this a confidential response? (select as appropriate)

No

To provide adequate moderation and redress mechanisms, providers should deploy trained human moderators, with training that includes but is not limited to:

- How to identify risks to child safety, with awareness of the service's policies, what constitutes a violation of published terms, and how to implement them effectively and fairly
- Knowledge of risks to different groups of children, including gender-based risks and addressing violence against women and girls
- When and how to intervene, including to whom (internal and external stakeholders e.g. law enforcement) to escalate or refer issues
- The full range of activity that is illegal or might be harmful to a child, or that may constitute child abuse, and how to identify such material and behaviour on the service
- Knowledge of the stages of child development, with awareness of the evolving capacities, vulnerabilities and behaviours of children as they grow
- Up-to-date training on the latest advice for e.g. from regulators and experts regarding staying safe online so child users can be directed towards it
- Access to support to protect and manage their mental health and wellbeing

Question 24: How do human moderators and automated systems work together, and what is their relative scale? How should services guard against automation bias?

Question 24: How do human moderators and automated systems work together, and what is their relative scale? How should services guard against automation bias?
Is this a confidential response? (select as appropriate) *[Please select]*

Question 25: In what instances is content that is harmful to children, that is in contravention of terms and conditions, removed from a service or the part of a service that children can access?
Is this a confidential response? (select as appropriate) *[Please select]*

Question 26: What other mitigations do services currently have to protect children from harmful content?
Is this a confidential response? (select as appropriate) *[Please select]*

Question 27: Where children attempt to circumvent mitigations in place on a service, what further systems and processes can a service put in place to protect children?
Is this a confidential response? (select as appropriate) *[Please select]*

Question 28: Other than those covered above in this document (the call for evidence), are you aware of other measures available for mitigating the risk, and impact of, harm from content that is harmful to children?
Is this a confidential response? (select as appropriate) *[Please select]*