Your response

Volume 2: The causes and impacts of online harm

Ofcom's Register of Risks

Question 1:
i) Do you have any comments on Ofcom's assessment of the causes and impacts of online harms?

WeProtect Global Alliance ('the Alliance') is a non-profit that brings together people and organisations with the knowledge, experience and influence to transform the global response to child sexual exploitation and abuse online. As of February 2024, its membership is comprised of 102 government members – including the Government of the United Kingdom – 70 companies, 94 civil society groups and 10 international organisations. As a multi-sector membership organisation spanning governments, civil society, the private sector and international non-governmental organisations, WeProtect Global Alliance occupies a unique position in the child protection sector and thus has a comprehensive viewpoint of both the global threat of child sexual exploitation and abuse online and the current response. Answers to this consultation will focus on this specific illegal online harm and connected harms.

Ofcom's assessment of the causes and impacts of online harms is comprehensive, clear and in line with the findings of the Alliance. Child sexual abuse content is a serious crime that can have devastating emotional, social and physical consequences for victims and survivors. The scale and complexity of technology-facilitated child sexual abuse is increasing rapidly. In 2022 alone, NCMEC received 32 million Reports of suspected child sexual abuse material. WeProtect Global Alliance's 2023 Global Threat Assessment found that known forms of child sexual exploitation and abuse online, such as 'self-generated' material, financial sexual extortion and grooming have intensified and evolved in recent years. The report also stresses that new trends such as AI-generated child sexual abuse material (CSAM), eXtended Reality (XR) and sexualised legal content are further fuelling offending. A recent study by the Alliance, conducted by Economist Impact, explored the experiences of 2,000 18-year-olds in Europe who had regular access to the internet as children. The study aimed to understand their experiences of and exposure to online sexual harms during childhood and found 68% of those surveyed reported experiencing at least one sexual harm online during childhood. The two most common types of harm experienced by European children receiving sexually explicit content from an adult they know or did not know (65%) and being asked to do something sexually explicit online that they did not want to do or were uncomfortable with (55%). Research conducted by the International Justice Mission and University of Nottingham Rights Lab found that in 2022 alone, approximately 1 in 100 children in the Philippines were trafficked to produce child sexual exploitation material, driven in demand by perpetrators in countries such as the United Kingdom.

Research conducted by the NSPCC indicates very real and lasting impacts resulting from child sexual exploitation and abuse online. It found that while there are often perceptions of online

sexual harms having less impact on children, in reality technology assisted child sexual abuse was stressed as being no less impactful than "offline abuse". In addition to this, the study showed that child sexual exploitation and abuse online adds additional challenges for children and young people, including control, permanence, blackmail, revictimisation and self-blame.

ii) Do you think we have missed anything important in our analysis? Please provide evidence to support your answer.

It is of paramount importance that we invest resources into identifying and better understanding the causes of child sexual exploitation and abuse, including risk and protective factors. The 2023 Global Threat Assessment identifies this as a crucial component of developing a public health response to child sexual exploitation and abuse, which have thus far tended to be under-prioritised and under-funded. Safety by Design is a crucial element of the preventative approach. While many of the measures proposed by Ofcom will ensure that digital services implement or upgrade safety features on their services, it is important to explicitly mention Safety by Design. Although Safety by Design principles were first introduced in 2018, there is still insufficient transparency to ascertain the extent to which Safety by Design is being implemented, and the effectiveness of measures adopted. According to an OECD report on Transparency Reporting on Child Sexual Exploitation and Abuse Material Online by the Global Top-50 Content Sharing Services, 30 of the top 50 online platforms do not issue transparency reports covering steps taken to combat child sexual exploitation and abuse online. Insights from the Tech Coalition's annual survey, administered with the WeProtect Global Alliance to 31 coalition member companies, also signal an increasing but relative lack of investment in measures associated with Safety by Design. While the adoption of technologies to detect existing illegal or harmful content at both network and platform level is advanced, the adoption of 'age oriented online safety' and 'user protection' is still in development, suggesting that companies are slower to adopt measures to prevent harm from occurring in the first place.

iii) Is this response confidential? (if yes, please specify which part(s) are confidential)

No

Question 2:

i) Do you have any views about our interpretation of the links between risk factors and different kinds of illegal harm? Please provide evidence to support your answer.

The Alliance aligns with Ofcom's interpretation of the links between risk factors and different kinds of illegal harm. The Alliance understands and agrees with the highlighted benefits of technological features such as end-to-end encryption (E2EE) and livestreaming, but also finds these technological features to play a significant role in the abuse and sexual exploitation of children in online environments.

While E2EE offers important privacy protections for children as much as adults, the Alliance's 2023 Global Threat Assessment demonstrates how many perpetrators who attempt to groom children online firstly identify targets on social media, in chat rooms, gaming environments before diverting conversations to a private messaging app or an end-to-end encrypted environment due to the lower risk of detection. This technique is known as 'off-platforming'. If not accompanied by the implementation of appropriate measures to mitigate risks to children (through a Safety by Design approach), there is a high risk that E2EE environments make it impossible for third parties

to detect child sexual exploitation and abuse online, thereby denying both platforms and law enforcement the opportunity to proactively detect, report, and remove child sexual abuse material. E2EE also hinders the visibility of the threat. NCMEC anticipates that with the widespread adoption of E2EE, the number of reports of suspected child sexual abuse from larger platforms (of which five accounted for 93% of all reports in 2022) will decrease by almost 80%. This likely drop in reporting is a significant concern. Although industry reports provide only a partial view of the scale of child sexual exploitation and abuse online, they are crucial to informing the global response, particularly in light of low levels of reporting by victim-survivors.

The Alliance's 2023 Global Threat Assessment also found that the scale of livestreamed child sexual abuse is difficult to ascertain for a number of interrelated reasons. First, livestreaming of child sexual abuse is not consistently criminalised around the world. Second, even in countries where it is an offence, livestreaming is often difficult to investigate and prosecute because once the livestream is over, there may be little evidence unless it was recorded. Third, most platforms don't monitor private livestreams. In August 2022, the Australian e-Safety Commissioner issued the first mandatory transparency notices to Microsoft, Skype, Snap, Apple, Meta, WhatsApp, and Omegle, four of which have livestreaming or video call/conferencing services. Responses revealed that of these four, three do not currently use tools to detect livestreamed child sexual abuse or exploitation. The 2021 Global Threat Assessment noted that COVID-19 related travel restrictions fuelled an increase in this type of abuse since offenders could not travel. Additionally, 39% of respondents to Suojellaan Lapsia's survey of dark web users reported they had viewed child sexual abuse livestreaming, indicating significant demand. The survey has broad reach but the findings may be more representative of the habits of offenders with a propensity to seek help, as respondents participated voluntarily.

The 2023 Global Threat Assessment also stresses further risk factors and emerging threats that should be highlighted in this response. As already highlighted in the summary note, generative AI tools are making it easier and quicker to produce illegal content rapidly and at scale. Open-source AI is already being used to produce results of infinite images by generating new images of abuse as well as modifying and distorting existing images of children. This includes increasingly violent and harmful versions of victims' abuse. Offenders are also using AI to share tips and generate demand, creating an ever-faster cycle of production. When real images are used to generate AI images such as sexual "deepfakes", the impact on those affected can be catastrophic. On top of this, the potential volume of new material threatens to overwhelm law enforcement agencies in identifying and rescuing at-risk children and bringing offenders to justice. As AI-generated content becomes increasingly realistic and convincing, the challenge for law enforcement and hotline services will be to differentiate between "synthetic" and real abuse material in order to save children from immediate danger, increasing the burden on already over-stretched services.

Emerging technologies like eXtended Reality (XR) pose additional risks for child safety. The first days of 2024 have seen British police begin to investigate the rape of a girl under the age of 16 in the metaverse, which is the first investigation of its kind. As set out in our joint intelligence briefing on XR technologies and child sexual exploitation and abuse with the University of Manchester, there is currently limited other evidence of use of XR in child sexual abuse exploitation and abuse. However, risks include opportunities for offenders to access victim-survivors; to distribute child sexual abuse material; simulate abuse of virtual representations of children; and use integrated tech such as haptics, which simulate real world sensations such as movements, vibrations, and force. While there are signs of slowing enthusiasm and investment in the metaverse, the general upwards trajectory remains undeniable. The global market for XR is

forecasted to surpass $1.1 trillion by 2030. It is likely offenders will increasingly exploit XR technologies as they become more accessible and affordable, given that detection of sexual exploitation and abuse in such environments remains challenging.
ii) Is this response confidential? (if yes, please specify which part(s) are confidential)
No

Volume 3: How should services assess the risk of online harms?

Governance and accountability

Question 3:
i) Do you agree with our proposals in relation to governance and accountability measures in the illegal content Codes of Practice?
N/A
ii) Do you think we have missed anything important in our analysis? Please provide evidence to support your answer.
N/A
iii) Is this response confidential? (if yes, please specify which part(s) are confidential)
No

Question 4:
i) Do you agree with the types of services that we propose the governance and accountability measures should apply to?
N/A
ii) Please explain your answer.
N/A
iii) Is this response confidential? (if yes, please specify which part(s) are confidential)
No

Question 5:
i) Are you aware of any additional evidence of the efficacy, costs and risks associated with a potential future measure to requiring services to have measures to mitigate and manage illegal content risks audited by an independent third-party?
N/A
ii) Is this response confidential? (if yes, please specify which part(s) are confidential)

No

Question 6:	
i)	Are you aware of any additional evidence of the efficacy, costs and risks associated with a potential future measure to tie remuneration for senior managers to positive online safety outcomes?
N/A	
ii)	Is this response confidential? (if yes, please specify which part(s) are confidential)
No	

Service's risk assessment

Question 7:	
i)	Do you agree with our proposals?
Yes	
ii)	Please provide the underlying arguments and evidence that support your views.

The Alliance is supportive of the outcomes-based approach outlined in the chapter summary paper by Ofcom. It is true that different services and technologies have different risk profiles, particularly when it comes to child sexual exploitation and abuse online. Illegal harmful content, where the severity and immediate impact of the harm is most dangerous, or that poses an imminent threat to the lives and safety of vulnerable users, should attract the most stringent risk mitigation measures. Child sexual abuse content is a serious crime that can have devastating emotional, social and physical consequences for victims and survivors.

Platforms and digital services which are most at risk of causing serious and immediate harm through facilitating the most abhorrent crimes, such as child exploitation and abuse online, should be prioritised. Size, impact and market share should also be considered when determining how to prioritise the regulation of such a wide variety of platforms. In addition to these criteria, there is also a strong case to be made for a higher level of scrutiny of platforms that are particularly popular with and regularly used by children. It is essential that such platforms adhere to the rules and provide products which provide a safe, positive experience for child users, based on a sophisticated understanding of likely risks. They have a unique and critical responsibility in ensuring services are safe by design, that their teams have the resources and infrastructure to quickly detect and tackle harm and provide support to their more vulnerable users.

Industry stakeholders should have proactive, robust and comprehensive risk mitigation measures in place to detect, report, block and remove this content (both new child sexual abuse material and material which has already been identified), as well as to report it to the authorities in order for offenders to be brought to justice. In order to be effective, the type and way in which the mitigation measures are implemented should be flexible and allow the private sector to adopt solutions based on the specific risks on their services. There will need to be explicit and strict safety obligations on high priority issues such as child sexual abuse and exploitation online, and for platforms with a larger market share or high-risk profile.

In explaining the risk mitigation measures adopted, tech companies should be able to provide evidence that user safety was prioritised in product design and engineering decisions and that they have adopted a Safety by Design approach.
iii) Is this response confidential? (if yes, please specify which part(s) are confidential)
No

Specifically, we would also appreciate evidence from regulated services on the following:

Question 8:
i) Do you think the four-step risk assessment process and the Risk Profiles are useful models to help services navigate and comply with their wider obligations under the Act?
Yes
ii) Please provide the underlying arguments and evidence that support your views.
The four-step risk assessment process to ((i) understand the harms that need to be assessed; (ii) assess risks by considering the likelihood and potential impact of harms occurring on their service; (iii) implement safety measures and record outcomes of the risk assessment; and (iv) report, review and update the risk assessment) is clear and aligned with our Model National Response framework. The ability to review and update risk assessments is particularly important given that the landscape of illegal online harms is a quick-moving and consistently changing landscape.
iii) Is this response confidential? (if yes, please specify which part(s) are confidential)
No

Question 9:
i) Are the Risk Profiles sufficiently clear?
Yes, but further, more granular clarification would be helpful
ii) Please provide the underlying arguments and evidence that support your views.
The Alliance is supportive of the risk-based approach that Ofcom is suggesting and strongly supportive of the recommendation for services to have a written policy in place to review their assessment at least every 12 months. User reports, user complaints, user data including age (where relevant), retrospective analysis of incidents of harm are all useful indicators that can help build a comprehensive picture of the risk profile of an online service or platform. It would be helpful to know specific the tables listing risk factors will be, for example, if it will be broad categories such as "child sexual exploitation and abuse online" or whether it will cover a targeted breakdown of each illegal online harm within each harm category such as "online grooming", "livestreaming child sexual abuse", "child sexual abuse material", "AI-generated child sexual abuse", "grooming manuals/guidance" and so forth. Each of these illegal harms require different responses from industry players and manifest differently on different services. It would therefore be preferable to provide categories that accurately reflect the diversity of the harm types in order to design the most effective and efficient response.
iii) Do you think the information provided on risk factors will help you understand the risks on your service?
Not applicable
iv) Please provide the underlying arguments and evidence that support your views.
Not applicable
v) Is this response confidential? (if yes, please specify which part(s) are confidential)
No

Record keeping and review guidance

Question 10:
i) Do you have any comments on our draft record keeping and review guidance?
The proposals are clear and easy to understand. Requiring online services to keep a record of their risk assessments and compliance measures (and therefore resulting product and policy changes) will not only be useful to increase transparency, but also to build knowledge and understanding about how the private sector is responding to online illegal harms. Over time, the ability to follow the changes to risk assessments and compliance measures will allow the regulator to identify good practice and practices that require improvement. While allowing services to choose their preferred medium for record-keeping is practical and accommodates diverse needs, specifying minimum durability standards for services ensures accountability, facilitates reviews and could prevent data loss.
ii) Please provide the underlying arguments and evidence that support your views.

N/A	
iii)	Is this response confidential? (if yes, please specify which part(s) are confidential)
No	

Question 11:	
i)	Do you agree with our proposal not to exercise our power to exempt specified descriptions of services from the record keeping and review duty for the moment?
N/A	
ii)	Please provide the underlying arguments and evidence that support your views.
N/A	
iii)	Is this response confidential? (if yes, please specify which part(s) are confidential)
No	

Volume 4: What should services do to mitigate the risk of online harms

Our approach to the Illegal content Codes of Practice

Question 12:	
i)	Do you have any comments on our overarching approach to developing our illegal content Codes of Practice?

The Secretariat of WeProtect Global Alliance is supportive of Ofcom's approach to developing Codes of Practice for illegal harmful content online. The focus on the core areas of the design, operation and use of digital services is critical to designing and implementing effective tailored responses to specific threats, including those which contribute to child sexual exploitation and abuse online.

We particularly welcome the targeted measures which aim to set standards in tackling particularly challenging illegal online harms and content, such as the proliferation of child sexual abuse material (CSAM) on U2U services and search services. A raft of effective content moderation measures that detect and remove child sexual exploitation and abuse online already exist and it is important that all digital platforms at risk of being exploited for this crime implement the effective, minimum standard tools in this space. Effective detection of 'known' child sexual abuse material is made possible by two linked techniques called 'hashing' and 'hash-matching'. These techniques have significantly accelerated the identification and removal of known child sexual abuse material from the internet. In addition to these techniques, the development of automated or semi-automated AI classifiers has been incredibly useful in the detection, reporting, removal and blocking of 'unknown' or 'new' child sexual abuse material online. In a 2021 survey of tech company practices, conducted by WeProtect Global Alliance and the Tech Coalition, 84% of the companies surveyed said they had at least partly automated processes for forwarding reports of child sexual abuse online, suggesting that report management is growing increasingly efficient.

Digital services should continue to work – in partnership with safety tech experts and industry – on enhancing the accuracy of classifiers to detect 'unknown' child sexual abuse content (including livestreamed content) and grooming in both non-encrypted and encrypted video sharing environments.

The focus on good practice within industry will allow us to collectively leverage existing knowledge, remain on the "digital front foot", foster collaboration and setting high standards. By sharing successful approaches, platforms can learn from each other, leading to more comprehensive solutions.

ii)	Is this response confidential? (if yes, please specify which part(s) are confidential)

No

Question 13:

i)	Do you agree that in general we should apply the most onerous measures in our Codes only to services which are large and/or medium or high risk?

Yes

ii)	Please provide the underlying arguments and evidence that support your views.

As above, different services and technologies have different risk profiles, particularly when it comes to child sexual exploitation and abuse online. Illegal harmful content, where the severity and immediate impact of the harm is most dangerous, or that poses an imminent threat to the lives and safety of vulnerable users, should attract the most stringent risk mitigation measures. Child sexual abuse content is a serious crime that can have devastating emotional, social and physical consequences for victims and survivors.

Platforms and digital services which are most at risk of causing serious and immediate harm through facilitating the most abhorrent crimes, such as child exploitation and abuse online, should be prioritised. Size, impact and market share should also be considered when determining how to prioritise the regulation of such a wide variety of platforms. In addition to these criteria, there is also a strong case to be made for greater scrutiny of platforms that are marketed to and/or are regularly used by children. It is essential that such platforms adhere to the rules and provide products which provide a safe, positive experience for child users, based on a sophisticated understanding of likely risks. They have a unique and critical responsibility in ensuring services are safe by design, that their teams have the resources and infrastructure to quickly detect and tackle harm and provide support to their more vulnerable users.

Industry stakeholders should have proactive, robust and comprehensive risk mitigation measures in place to detect, report, block and remove this content (both new child sexual abuse material and material which has already been identified), as well as to report it to the authorities in order for offenders to be brought to justice. In order to be effective, the type and way in which the mitigation measures are implemented should be flexible and allow the private sector to adopt solutions based on the specific risks on their services. There will need to be explicit and strict safety obligations on high priority issues such as child sexual abuse and exploitation online, and for platforms with a larger market share or high-risk profile.

In explaining the risk mitigation measures adopted, tech companies should be able to show that user safety was prioritised in product design and engineering decisions and that they have adopted a Safety by Design approach.
iii) Is this response confidential? (if yes, please specify which part(s) are confidential)
No

Question 14:
i) Do you agree with our definition of large services?
Yes
ii) Please provide the underlying arguments and evidence that support your views.
WeProtect Global Alliance is glad to see that the definition of large services closely mirrors that of the European Union's Digital Services Act, where very large online platforms (VLOPs) and very large online search engines (VLOSEs) are defined as having 45 million active EU users per month, which accounts for approximately 10% of the EU population. Given the transnational nature of the internet, international alignment on standards and rules is an important part of the Global Strategic Response to tackle illegal online harms, such as online child sexual exploitation and abuse. Not only does it ensure consistency in terms of the requirements and expectations for platforms and digital services in responding to online illegal harms, but it also reduces the burden on industry and facilitates a more effective and efficient response, reducing administrative burden.
iii) Is this response confidential? (if yes, please specify which part(s) are confidential)
No

Question 15:	
i)	Do you agree with our definition of multi-risk services?
Yes	
ii)	Please provide the underlying arguments and evidence that support your views.
The Alliance agrees that with the definition of a service as being multi-risk where it is found to be high or medium risk for at least two kinds of illegal harms.	
iii)	Is this response confidential? (if yes, please specify which part(s) are confidential)
No	

Question 16:	
i)	Do you have any comments on the draft Codes of Practice themselves?
Ofcom's proposed Codes of Practice represent a positive step toward protecting children online by tackling illegal content. They offer a much-needed framework for online platforms to address harmful and illegal material, promoting safer digital spaces for young users. However, it's crucial to acknowledge that the evolving nature of online threats necessitates further development to effectively address emerging dangers. They establish clear expectations for platforms regarding content moderation, risk assessment, and user reporting, fostering accountability and transparency. The Codes also provide a consistent framework across platforms and clearly prioritise user safety by emphasising swift removal of illegal content and providing reporting mechanisms for users. WeProtect Global Alliance would like to see specific measures for tackling emerging and evolving threats surrounding child sexual exploitation and abuse online, such as AI-generated illegal content (including deepfakes or manipulated images) and illegal content in Extended Reality (XR) environments (where immersive experiences can heighten the impact of harm). Generative AI tools have become mainstream in a short period of time. While calling for remedies for specific harms, the Codes should also be adaptable and regularly reviewed to address evolving challenges.	
ii)	Is this response confidential? (if yes, please specify which part(s) are confidential)
No	

Question 17:	
i)	Do you have any comments on the costs assumptions set out in Annex 14, which we used for calculating the costs of various measures?
No	
ii)	Is this response confidential? (if yes, please specify which part(s) are confidential)
No	

Content moderation (User to User)

Question 18:	
i)	Do you agree with our proposals?
Yes	
ii)	Please provide the underlying arguments and evidence that support your views.
WeProtect Global Alliance is happy to see all digital services will be required to implement content moderation processes to take down illegal content quickly and effectively.	
iii)	Is this response confidential? (if yes, please specify which part(s) are confidential)
No	

Content moderation (Search)

Question 19:	
i)	Do you agree with our proposals?
Yes	
ii)	Please provide the underlying arguments and evidence that support your views.
WeProtect Global Alliance agrees that all search services should have systems and processes that are designed so that search results that is illegal content is deprioritised or deindexed for UK users.	
iii)	Is this response confidential? (if yes, please specify which part(s) are confidential)
No	

Automated content moderation (User to User)

Question 20:	
i)	Do you agree with our proposals?
Partly	
ii)	Please provide the underlying arguments and evidence that support your views.

The Alliance is supportive of the proposal for both smaller and larger services that have a specific risk of hosting or facilitating child sexual exploitation and abuse online to deploy hash matching technology across their services. Automated content moderation tools are essential in the detection, reporting, blocking and removal of child sexual abuse material (CSAM) and URLs which have been previously identified as hosting CSAM. Automated systems are needed since the sheer volume of such illegal content online – and the ever-increasing amount of child abuse online – would be almost impossible to tackle at scale without them. The Alliance agrees that U2U services that are at medium and high risk of being used to disseminate CSAM or be used for the sexual exploitation of children should be covered by the code. The monthly user threshold of 700,000 monthly users in the UK seems proportionate. The Alliance is particularly supportive that file-storage and file-sharing sites will be covered by the code. Until now it has been unclear to what extent detection tools have been deployed on file-storage and file-sharing sites and as the Alliance's 2021 Global Threat Assessment showed, perpetrators of child sexual abuse typically use cloud file sharing to efficiently exchange images and videos with both known and new offender contacts. To ensure that content remains accessible for as long as possible, determined offenders use multiple cloud platforms simultaneously. The true nature of harmful links is hidden behind a smoke screen of references to other (lesser) illegal activity or legitimate file-sharing uses to evade detection.

In addition to hash matching, other forms of detecting child sexual exploitation and abuse online should be encouraged. The development of AI classifiers has been incredibly useful in the detection, reporting, removal and blocking of 'unknown' or 'new' child sexual abuse material online and could potentially play a vital role in tackling emerging threats such as AI generated material, livestreaming and abuse in extended reality environments. Such automated or semi-automated moderation systems identify illegal content by following rules and interpreting many

different examples of content which is and is not illegal. In recent years, efforts have focused on improving classifiers to detect 'new' child sexual abuse imagery. Video classifiers are the least mature and have lower accuracy rates, making automated triage difficult. Without innovation to develop new solutions, this could significantly increase pressure on moderators and analysts given the greater proportion of video imagery that is being reported. Of the Cyber Tipline reports received by NCMEC in 2022, 88 million contained media files, of which 37 million were videos. Many existing classifiers are underpinned by powerful machine learning and AI capabilities, which are improving at pace.

The Secretariat of the Alliance has some concerns when it comes to the scope of the measures. For example, automated systems are not recommended for private communications and end-to-end encrypted environments. The 2023 Global Threat Assessment highlights how many perpetrators prioritise anonymity, ease of access, and availability of material or victim-survivors when considering how and where to conduct child sexual abuse online, it is likely that the use of E2EE group messaging services is likely to grow in the future. The 2023 Global Threat Assessment also identified a number of ways in which automated detection tools can be implemented in E2EE environments, including client-side scanning (which involves scanning messages on devices for matches or similarities to a database of illegal child sexual abuse material before the message is encrypted and sent), homomorphic encryption (the use of a different type of encryption which allows operations to be performed without data decryption at any point) and intermediate secure enclaves (which decrypt the message at server level by a third party and use tools to detect child sexual abuse materials). WeProtect Global Alliance supports the future-proof approach of the Online Safety Act which cover all types of threats and adopts a tech neutral approach. If the proposed measures do not cover private communications and E2EE environments, it would be useful to hear proposals on how illegal online harm should be tackled on such services.

Is this response confidential? (if yes, please specify which part(s) are confidential)

No

Question 21:

Do you have any comments on the draft guidance set out in Annex 9 regarding whether content is communicated 'publicly' or 'privately'?

We note that end to end encrypted messaging services and other forms of "private communications" continue to be preferred by many perpetrators due to the lower perceived risk of detection. Given that many perpetrators prioritise anonymity, ease of access, and availability of material or victim-survivors when considering how and where to conduct child sexual abuse online,220 it is likely that the use of E2EE group messaging services is likely to grow in the future. This could lead to wider distribution of child sexual abuse material, and the sharing of techniques to perpetrate abuse and evade law enforcement in the absence of other safeguards. Ofcom should be mindful of these risks and ensure they are considered in the overall implementation of the Online Safety Act.

Is this response confidential? (if yes, please specify which part(s) are confidential)

No

Do you have any relevant evidence on:

Question 22:
Accuracy of perceptual hash matching and the costs of applying CSAM hash matching to smaller services;
Accuracy of perceptual hash matching should be regularly reviewed to ensure it is consistent with current best practice and reflects the development of relevant tools. It is important to consider the costs of applying CSAM hash matching tools to smaller services. While there are a number of hash matching technologies that are available to services free of charge (for example Microsoft's PhotoDNA Cloud Service is "a free service for qualified customers and developers" and Google's CSAI Match is licensed "to a number of other technology companies free of charge"), the acquisition, deployment and management of such systems will require headcount.
Please provide the underlying arguments and evidence that support your views.
See above
Is this response confidential? (if yes, please specify which part(s) are confidential)
No

Question 23:
Ability of services in scope of the CSAM hash matching measure to access hash databases/services, with respect to access criteria or requirements set by database and/or hash matching service providers;
N/A
Please provide the underlying arguments and evidence that support your views.
N/A
Is this response confidential? (if yes, please specify which part(s) are confidential)
No

Question 24:
Costs of applying our CSAM URL detection measure to smaller services, and the effectiveness of fuzzy matching for CSAM URL detection;
WeProtect Global Alliance's 2021 Global Threat Assessment highlighted that there are signs of offenders moving away from the curation of personal collections and preferring 'on-demand' access to content via the sharing of links that lead to child sexual abuse content. Links to files containing child sexual abuse content are posted across multiple sites and often used as part of offender-to-offender sharing. This creates a raft of challenges for law enforcement. Material is often published and hosted in different jurisdictions, which complicates evidence-gathering. The volume of content in an offender's possession was historically one of several factors used to assess the level of risk they posed, but this is no longer always indicative. A report by WeProtect

Global Alliance – in partnership with GCHQ – highlights how different industry players are currently responding to the threat of link sharing in a number of ways and to varying degrees.

There is little comparable data available on how companies are responding, which makes it difficult to assess the efficacy of responses, as shown by the OECD's 2023 Report on Transparency Reporting on Child Sexual Abuse and Exploitation Online. A challenge for many service providers is how to moderate links on their platform where the content is hosted on a different site. The action taken by industry can depend on where the links take users. For example, a link may take a user to content hosted externally, or link to an image-hosting site or website, or to group chats on group messaging apps and forums. All these may be harmful yet require differing responses. Collaboration with leading safety technology organisations forms an essential part of the response for leading industry players. The Internet Watch Foundation's (IWF) URL List is cited as a helpful tool in identifying potential harms and blocking access to illicit webpages and material. The IWF is constantly updating and reviewing this list – twice a day – for its members to use in tackling the dissemination of links to child sexual abuse online. In addition to ensuring that access to child sexual abuse material is blocked, the IWF works with relevant actors to ensure that the images and videos at the linked location are removed from the internet. Project Arachnid in Canada is also highlighted as an effective technology to combat link-sharing. Project Arachnid identifies child sexual abuse material by crawling specific publicly accessible URLs reported to CyberTipline, as well as URLs on the surface web3and dark web4 that are proven or known to host child sexual abuse material. It detects URLs that host media and matches content against a database of digital fingerprints. As soon as Project Arachnid detects a match in fingerprints, a removal notice is automatically issued requesting the hosting provider to take it down. It follows up on this request by recrawling URLs linking illegal content every day until the content is taken down.

As noted above, sharing of trust and safety technologies and more consistent transparency reporting, policies and processes will be essential in facilitating a more effective response. Constant technological innovation and the increased deployment of artificial intelligence will be required to respond to the scale and complexity of the threat.

i)	Please provide the underlying arguments and evidence that support your views.
See above	
ii)	Is this response confidential? (if yes, please specify which part(s) are confidential)
No	

Question 25:	
i)	Costs of applying our articles for use in frauds (standard keyword detection) measure, including for smaller services;
Not comment	
ii)	Please provide the underlying arguments and evidence that support your views.
Not applicable	
iii)	Is this response confidential? (if yes, please specify which part(s) are confidential)
No	

Question 26:	
i)	An effective application of hash matching and/or URL detection for terrorism content, including how such measures could address concerns around 'context' and freedom of expression, and any information you have on the costs and efficacy of applying hash matching and URL detection for terrorism content to a range of services.
No comment	
ii)	Please provide the underlying arguments and evidence that support your views.
Not applicable	
iii)	Is this response confidential? (if yes, please specify which part(s) are confidential)
No	

Automated content moderation (Search)

Question 27:	
i)	Do you agree with our proposals?
Yes	
ii)	Please provide the underlying arguments and evidence that support your views.
The Alliance agrees with the proposal to require both small and large search services at low, medium and high risk to deindex URLs which have been previously identified as hosting CSAM or which include a domain identified as dedicated to CSAM from the search index. Deindexing CSAM-linked URLs offers a multifaceted approach to protecting children. It aims to reduce accessibility, disrupt offender behaviour and ease the load on law enforcement. Additionally, it sends a strong message against easy CSAM distribution/access and requires platforms to play their role in reducing its normalisation. We would welcome further clarification on why the measure is recommended for general search services only and does not include vertical search services.	
iii)	Is this response confidential? (if yes, please specify which part(s) are confidential)
No	

User reporting and complaints (U2U and search)

Question 28:	
i)	Do you agree with our proposals?
Yes, however the onus on addressing and remedying child sexual abuse and exploitation should not be on child users of platforms or services, nor on parents and caregivers.	
ii)	Please provide the underlying arguments and evidence that support your views.
The Alliance believes it is particularly important that all relevant platforms and services have an easy to find, easy to access and easy to use complaints system. We also welcome the direct reference to particularly vulnerable users or those with accessibility needs. Our Global Threat	

Assessment has highlighted disabled children as being particularly vulnerable to child sexual abuse and exploitation, and so it is critical that they can access clear and tailored reporting mechanisms. However, data from hotlines such as IWF and NCMEC demonstrates that the vast majority of their reports come from industry rather than user reports. While clear and accessible user reporting is essential, it will not substitute for effective and proactive safety by design measures in relevant platforms and services.
iii) Is this response confidential? (if yes, please specify which part(s) are confidential)
No

Terms of service and Publicly Available Statements

Question 29:	
i)	Do you agree with our proposals?
Yes	
ii)	Please provide the underlying arguments and evidence that support your views.
Regarding child sexual abuse content specifically, the terms of services for all digital services should clearly state that the platform has a zero-tolerance approach regarding child sexual exploitation and abuse on its services. Terms and conditions should be concise and clear in defining what child sexual abuse material entails (photographs, videos, live streaming, grooming and digital or computer generated images, including the current emerging threat of AI-generated content), that such material is prohibited, how users can report child sexual abuse material and what the consequences will be for posting such content (ban from platform, account deletion, referral to law enforcement, investigation and possible prosecution). For platforms that are marketed to and/or regularly used by children, terms of service and publicly available statements should be provided in versions accessible to child users.	
iii)	Is this response confidential? (if yes, please specify which part(s) are confidential)
No	

Question 30:	
i)	Do you have any evidence, in particular on the use of prompts, to guide further work in this area?
N/A	
ii)	Please provide the underlying arguments and evidence that support your views.
N/A	
iii)	Is this response confidential? (if yes, please specify which part(s) are confidential)
No	

Default settings and user support for child users (U2U)

Question 31:	
i)	Do you agree with our proposals?
Yes	
ii)	Please provide the underlying arguments and evidence that support your views.
The Alliance welcomes the specific measures that will apply to all users under the age of 18. The Alliance understands that, for the time being, these would only apply to the extent that a service has an existing means of identifying child users and would apply where the information available to services indicates that a user is a child, but calls for a more comprehensive approach to the	

implementation of age assurance technologies on U2U services. Just as we protect children offline – they can't freely walk into a nightclub or buy a bottle of wine – the same protections need to be implemented online. Whilst there are many positive opportunities available online, increasing numbers of children are accessing explicit content, chatting to strangers or being coerced into sharing images of themselves. Research shows that many children - some as young as 7 years old - stumble upon adult pornography online, with 61% of 11-13-year-olds describing their viewing as mostly unintentional. WeProtect Global Alliance believes that age assurance is one of the tools that can be used to create digital products safe by design. Our 2021 Global Threat Assessment highlighted that age estimation and verification tools are some of the Safety by Design solutions with the most potential to reduce the risk of online grooming. Such technology is still relatively nascent but could be used to exclude predators from children's forums and ensure age-appropriate online experiences. There are many different methods for carrying out age assurance checks, from more 'traditional' types such as ID, mobile phone number or credit card checks, to evolving technologies such as facial age estimation, identity apps and social media proofing. To ensure that users remain in control of their privacy, the Alliance believes that it is important to provide consumers with a choice as to which age estimation tools they use to confirm their age online. More information is available in our briefing on age estimation techniques.

The Alliance supports the clear guidance regarding improving the standards of default settings for child users of both smaller and larger U2U services where there is a specific risk of harm of grooming children for the purposes of sexual exploitation and abuse (CSEA). Through limiting network expansion prompts, hiding children's accounts from connection lists, blocking the ability of unknown accounts to send direct messages to children's accounts, implementing measures to ensure children do not receive unsolicited direct messages, keeping the location of children's accounts hidden it will make it hard for those seeking to groom, sexually exploit and abuse to get in contact with children.

The Alliance is also supportive of the suggested measures for smaller and larger U2U services to improve the provision of supportive information to children. Information, safeguarding, and support on services need to be accessible and easy to use for all users, especially specific groups, such as making tools child-friendly and disability friendly. While it is good to inform child users of the risks when they are seeking to disable the default settings, we should also be cautious that not too much onus is placed on children in taking these decisions. Platforms still have a responsibility to ensure that their tools and products are safe by design and that offenders and potential offenders are prevented from contacting children in the first place.

iii)	Is this response confidential? (if yes, please specify which part(s) are confidential)
No	

Question 32:	
i)	Are there functionalities outside of the ones listed in our proposals, that should explicitly inform users around changing default settings?
N/A	
ii)	Is this response confidential? (if yes, please specify which part(s) are confidential)
No	

Question 33:	
i)	Are there other points within the user journey where under 18s should be informed of the risk of illegal content?
There are several key points within the user journey where informing under-18s about the risk of illegal content is crucial. Firstly, before account creation. The terms of service and privacy policies should clearly state that illegal content is prohibited and outline potential consequences for accessing it. Simple, age-appropriate language is essential to communicate this. If such methods are used by platforms, during the age verification process, a clear message can be displayed highlighting what to do in case of illegal content, harmful behaviour, and potential risks. While using digital services algorithms should be designed to minimise the recommendation of illegal content to users under 18. Transparency about how recommendations are generated can also be helpful. When content is removed due to illegality, clear and informative messages explaining the reason and potential risks should be shown. Easy-to-find and understand reporting tools should be available for users to flag illegal content, although, as mentioned earlier in this consultation feedback, we should also be cautious with flagging that not too much onus is placed on users to report. Platforms have a responsibility to deploy systems that detect illegal content online and hire enough content moderators, provide training, and support them to ensure that they have the tools and resources to act swifty to identify and takedown harmful content. It's important to remember that there's no one-size-fits-all approach, and the most effective strategy will likely involve a combination of these different methods. Ultimately, the goal is to create a safe and informative online environment for all users, particularly those under 18, by proactively addressing the risk of illegal content.	
ii)	Is this response confidential? (if yes, please specify which part(s) are confidential)
No	

Recommender system testing (U2U)

Question 34:	
i)	Do you agree with our proposals?
N/A	
ii)	Please provide the underlying arguments and evidence that support your views.
N/A	
iii)	Is this response confidential? (if yes, please specify which part(s) are confidential)
No	

Question 35:	
i)	What evaluation methods might be suitable for smaller services that do not have the capacity to perform on-platform testing?
N/A	

ii)	Is this response confidential? (if yes, please specify which part(s) are confidential)
No	

We are aware of design features and parameters that can be used in recommender system to minimise the distribution of illegal content, e.g. ensuring content/network balance and low/neutral weightings on content labelled as sensitive.

Question 36:	
i)	Are you aware of any other design parameters and choices that are proven to improve user safety?
N/A	
ii)	Is this response confidential? (if yes, please specify which part(s) are confidential)
No	

Enhanced user control (U2U)

Question 37:	
i)	Do you agree with our proposals?
N/A	
ii)	Please provide the underlying arguments and evidence that support your views.
N/A	
iii)	Is this response confidential? (if yes, please specify which part(s) are confidential)
No	

Question 38:	
i)	Do you think the first two proposed measures should include requirements for how these controls are made known to users?
N/A	
ii)	Is this response confidential? (if yes, please specify which part(s) are confidential)
No	

Question 39:	
i)	Do you think there are situations where the labelling of accounts through voluntary verification schemes has particular value or risks?
N/A	
ii)	Is this response confidential? (if yes, please specify which part(s) are confidential)
No	

User access to services (U2U)

Question 40:	
i)	Do you agree with our proposals?
N/A	
ii)	Please provide the underlying arguments and evidence that support your views.
N/A	
iii)	Is this response confidential? (if yes, please specify which part(s) are confidential)
No	

Do you have any supporting information and evidence to inform any recommendations we may make on blocking sharers of CSAM content? Specifically:

Question 41:	
i)	What are the options available to block and prevent a user from returning to a service (e.g. blocking by username, email or IP address, or a combination of factors)?
There are several options available to block and prevent a user from returning to a service: • Blocking by username • Blocking by email • Blocking by IP address • Blocking by device fingerprint • Combined methods of blocking	
ii)	What are the advantages and disadvantages of the different options, including any potential impact on other users?
There are several important factors to consider when exploring the blocking of perpetrators who share child sexual abuse material online. Firstly, the chosen method should be appropriate for the reason behind blocking the user (e.g., violation of terms, security concerns). Blocking users should comply with relevant laws and ethical guidelines and users should be informed about the blocking process and have the opportunity to appeal. Each option available to block and prevent a user from returning to a service comes with advantages and limitations. • Blocking by username: while this is generally easy to implement and directly identifies the user at the source of the illegal content, blocking the username does not stop them coming back to the service and creating a a new account with a different username. • Blocking by email: this is generally considered more difficult for users to bypass and helps prevent offenders from creating new accounts with same email. However, users can also set up email addresses or secondary accounts to evade disruption. • Blocking by IP address: blocking access at the IP address level means that all access from a specific location or device is blocked. However, this is not very effective since users can very easily change IP addresses (usually by using a virtual private network (VPN)). At the same time, there is a risk that users who are using the internet perfectly legally at the same IP address are blocked access to the internet, for example, those using the same public Wi-Fi networks.	

- Blocking by device fingerprint: this is more sophisticated and uses unique device characteristics to identify users. Device fingerprints use a combination of hardware (operating system, model of device, serial numbers, etc.), software, and behavioural data to uniquely identify users, but privacy concerns and regulations exist. Despite being more complex, users can take steps to mitigate tracking, such as using privacy-focused browsers and tools like anti-fingerprinting extensions. Device fingerprinting can be privacy invasive.
- Combined methods of blocking: by combining multiple methods of blocking, the layering different blocking techniques can lead to a more robust type of blocking that disrupts offender behaviour in multiple ways. However, the multiple layers also increase complexity.

Ultimately, the best approach depends on the specific needs of the service and the type of users you are trying to block. It's important to weigh the effectiveness of each method against its potential drawbacks before implementing it.

iii)	Is this response confidential? (if yes, please specify which part(s) are confidential)

No

Question 42:
i) How long should a user be blocked for sharing known CSAM, and should the period vary depending on the nature of the offence committed?
Due to the illegal nature of the harm, we would support strict but proportionate sanctions in line with UK criminal law. Consideration should be given to an appropriate and proportionate response to accidental or "non malicious" sharing of known CSAM (e.g. inadvertently sharing while seeking advice on how to report).
ii) Is this response confidential? (if yes, please specify which part(s) are confidential)
No

There is a risk that lawful content is erroneously classified as CSAM by automated systems, which may impact on the rights of law-abiding users.

Question 43:
i) What steps can services take to manage this risk? For example, are there alternative options to immediate blocking (such as a strikes system) that might help mitigate some of the risks and impacts on user rights?
There is a risk of lawful content being misclassified as child sexual abuse material by automated systems, leading to potential infringement on user rights but there are also some key elements to help digital services avoid and remedy misclassification. The most important of which is to always implement a human review step in the process. By ensuring that flagged content is reviewed by trained human moderators before taking action, it allows for nuanced judgment and reduces false positives. It is the responsibility of those developing and deploying automated systems to detect child sexual abuse material to consistently invest in developing more sophisticated AI models that can better

distinguish between CSAM and lawful content. This might involve incorporating context, understanding artistic merit, and considering user intent. Content moderation policies and the decisions made based on them need to be clearly communicated. Users should be provided with avenues to appeal misclassifications.
ii) Is this response confidential? (if yes, please specify which part(s) are confidential)
No

Service design and user support (Search)

Question 44:	
i)	Do you agree with our proposals?
Yes	
ii)	Please provide the underlying arguments and evidence that support your views.
The Alliance supports the proposal that search requests where the wording clearly indicates that the user may be seeking to encounter Child Sexual Abuse Material (CSAM) and which use terms that explicitly relate to CSAM should provide content warnings and support resources to users. This form of secondary prevention is a key part of a preventative approach to reducing child sexual exploitation and abuse online. The 2023 Global Threat Assessment underscores that secondary prevention initiatives are typically focused on individuals at higher risk of violence perpetration. A study assessing the contribution of Stop it Now! Helplines in the UK, Ireland and the Netherlands shows they "can provide cost effective, quality advice and support [...] to prompt behaviour change in adults and strengthen protective factors which can reduce the risk of offending". Use of such support services is also increasing. According to the Lucy Faithfull Foundation, the number of people seeking advice or support via online self-help, or their confidential helpline has trebled since 2020. While it is positive that all large services are in scope of this proposal, we would welcome all digital services to invest more in a preventative response by disrupting searches on their services and signposting those at risk of offending to appropriate support services.	
iii)	Is this response confidential? (if yes, please specify which part(s) are confidential)
No	

Cumulative Assessment

Question 45:	
i)	Do you agree that the overall burden of our measures on low risk small and micro businesses is proportionate?
N/A	
ii)	Please provide the underlying arguments and evidence that support your views.
N/A	
iii)	Is this response confidential? (if yes, please specify which part(s) are confidential)
No	

Question 46:	
i)	Do you agree that the overall burden is proportionate for those small and micro businesses that find they have significant risks of illegal content and for whom we propose to recommend more measures?
N/A	

ii)	Please provide the underlying arguments and evidence that support your views.
N/A	
iii)	Is this response confidential? (if yes, please specify which part(s) are confidential)
No	

Question 47:	
i)	We are applying more measures to large services. Do you agree that the overall burden on large services proportionate?
Yes	
ii)	Please provide the underlying arguments and evidence that support your views.
As above, platforms and digital services which are most at risk of causing serious and immediate harm through facilitating the most abhorrent crimes, such as child exploitation and abuse online, should be prioritised. Size, impact and market share should also be considered when determining how to prioritise the regulation of such a wide variety of platforms. In addition to these criteria, there is also a strong case to be made for a higher level of scrutiny of platforms that are marketed to and/or regularly used by children. It is essential that such platforms adhere to the rules and provide products which provide a safe, positive experience for child users, based on a sophisticated understanding of likely risks. Larger services with a broader, global user base have a responsibility to ensure services are safe by design, that their teams have the resources and infrastructure to quickly detect and tackle harm and provide support to their more vulnerable users.	
iii)	Is this response confidential? (if yes, please specify which part(s) are confidential)
No	

Statutory Tests

Question 48:	
i)	Do you agree that Ofcom's proposed recommendations for the Codes are appropriate in the light of the matters to which Ofcom must have regard?
N/A	
ii)	Please provide the underlying arguments and evidence that support your views.
N/A	
iii)	Is this response confidential? (if yes, please specify which part(s) are confidential)
No	

Volume 5: How to judge whether content is illegal or not?

The Illegal Content Judgements Guidance (ICJG)

Question 49:	
i)	Do you agree with our proposals, including the detail of the drafting?
N/A	
ii)	What are the underlying arguments and evidence that inform your view?
N/A	
iii)	Is this response confidential? (if yes, please specify which part(s) are confidential)
No	

Question 50:	
i)	Do you consider the guidance to be sufficiently accessible, particularly for services with limited access to legal expertise?
N/A	
ii)	Please provide the underlying arguments and evidence that support your views.
N/A	
iii)	Is this response confidential? (if yes, please specify which part(s) are confidential)
No	

Question 51:	
i)	What do you think of our assessment of what information is reasonably available and relevant to illegal content judgements?
N/A	
ii)	Is this response confidential? (if yes, please specify which part(s) are confidential)
No	

Volume 6: Information gathering and enforcement powers, and approach to supervision.

Information powers

Question 52:	
i)	Do you have any comments on our proposed approach to information gathering powers under the Online Safety Act?

N/A	
ii)	Please provide the underlying arguments and evidence that support your views.
N/A	
iii)	Is this response confidential? (if yes, please specify which part(s) are confidential)
No	

Enforcement powers

Question 53:	
i)	Do you have any comments on our draft Online Safety Enforcement Guidance?
N/A	
ii)	Please provide the underlying arguments and evidence that support your views.
N/A	
iii)	Is this response confidential? (if yes, please specify which part(s) are confidential)
No	

Annex 13: Impact Assessments

Question 54:	
i)	Do you agree that our proposals as set out in Chapter 16 (reporting and complaints), and Chapter 10 and Annex 6 (record keeping) are likely to have positive, or more positive impacts on opportunities to use Welsh and treating Welsh no less favourably than English?
N/A	
ii)	If you disagree, please explain why, including how you consider these proposals could be revised to have positive effects or more positive effects, or no adverse effects or fewer adverse effects on opportunities to use Welsh and treating Welsh no less favourably than English.
N/A	
iii)	Is this response confidential? (if yes, please specify which part(s) are confidential)
No	