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?
Response: Overall, the assessment is a welcome review the varying issues associated with illegal harm occurring online. Our response to the next question identifies areas we believe were missed or ought to be developed more, with a focus on section 6C. Child Sexual Exploitation and Abuse.
ii) Do you think we have missed anything important in our analysis? Please provide evidence to support your answer.
Response: Below are the areas we believe were missed or ought be expanded upon. **5. Evidence and methodology** [redacted] **Reverse-image search and exposure of personal information**: Special considerations may apply to search services that provide reverse-image search functionality. If the initial image is of a survivor of CSAM, the function may return more CSAM images or pictures of the survivor from other context. If the service gatekeeps CSAM, a reverse-image search may still return legal images of survivors of CSAM or otherwise facilitate the identification of a survivor of online child sexual abuse or exploitation. As mentioned above, there are serious safety and privacy concerns whenever a search service can be used to find out information about a victim of CSAM. **6C.1 – 6C.106 - Grooming**

6C.106 – Revenue models: This section states there is no evidence revenues models are a risk factor in facilitating online grooming offences. We believe information in 6C.143 and 6W.9 is also relevant to grooming and should be included in 6C. Services that are low-capacity, at an early stage, or have a fast-growing user base, especially of children and youth, may be targeted by perpetrators. Children are often early adopters of new apps that may not have monetized their services yet or at least not fully, but this does mean there are not risky. One example is Wizz, discussed more below. Wizz has in-app purchases that enable users to purchase coins that can be used to add more friends or upgrade to a premium subscription. Another example is the Omegle website, which was a high-risk platform that had limited revenue-generating activities in the beginning, especially in comparison to some social medial companies.

Some platforms generate revenue through the sale of online gifts or tokens that can be transferred from one user to another. Some individuals who perpetrate online grooming offences send gifts/tokens to children as part of the grooming process and/or to incentivize the child to supply sexual imagery, making this form of revenue-generation for the company a risk factor for online grooming.

Finally, platforms that facilitate money transfers or other user-to-user payments may be attractive for those seeking to perpetrate financial sextortion schemes. Many of these platforms generate revenue from each transaction.

6C.107 – 6C.195 - Child Sexual Abuse Material

[redacted]

6C.140-141 – Adult Services: The inclusion of these paragraphs is important. The paragraphs mention C3P's Project Arachnid report, but we first raised in the issue of adolescent CSAM on adult pornography sites in our 2019 report *How We Are Failing Children: Changing the Paradigm* (see page 22). In terms of impact, we continue to see adolescents left without protection. There is no are barriers to upload, yet when a teen wants imagery down, we have examples of where they are asked for identification to prove they are underage in the image/video. This issue arises because in post-pubescence, the child has often reached full sexual maturity before the age of 18. Generally, females appear sexually mature at 12.5-18 years of age; the average for Caucasian females is 14-15 years old, and 13 years of age for African American females.

6C.143 – User base size: We echo the statements in this paragraph about perpetrators targeting smaller, less mature services. In our response to Question 8.1 we provide an example of a new service, which was renting servers from a Canadian company, that was quickly targeted by

perpetrators who used it to share CSAM. While the Canadian company had a policy for copyright violations, it had no policy for addressing CSAM and ignored over 200 notices that one its clients was being used to upload and share CSAM.

6C.186 – Generative AI: We are pleased to see this included and echo the statement that deepfake CSAM poses great concern in terms of detection and triaging content for victim identification purposes. In Canada, there is one reported legal decision on deepfake CSAM: *R v Larouche*, 2023 QCCQ 1853 (https://canlii.ca/t/k28rp). In this decision, the offender produced 7 deepfake CSAM videos, which equated to over 86,000 new CSAM images. The use of deepfake software to produce imagery was only detected because the investigating police offers were familiar with the appearance of certain victims and noticed certain videos in the collection looked off. The court noted how this technology could be used to victimize any child using photos stolen from social media or taken surreptitiously in public (see paragraph 70). It also stated that the presence of deepfake CSAM will have a major impact on community safety because of the influx of new CSAM with unknown hash values (paragraph 69).

We also direct Ofcom to a study that helps to explain the potential for realistic CSAM generated through AI. Near the end of 2023, the Stanford Internet Observatory released a report documenting its analysis of the dataset used in one of the most popular text-to-image GenAI tools, Stable Diffusion. The report concluded the dataset contained thousands of illegal child sexual abuse images. C3P helped to validate the findings of the study, which revealed through PhotoDNA analysis that over 3,000 pieces of suspected CSAM were found in a "small slice" of the billions of images in the data training set. Stanford's analysis noted that dataset has aided the creation of AI-generated CSAM (see Alexandra Levine, "Stable Diffusion 1.5 Was Trained On Illegal Child Sexual Abuse Material, Study Says, https://www.forbes.com/sites/alexandralevine/2023/12/20/stable-diffusion-child-sexual-abuse-material-stanford-internet-observatory/?sh=7a5cdcee5f21).

There are references to "GenAI chatbots" in the Annex to Volume 2, namely section **6W.21**. GenAI chatbots also have potential to be used to generate interactions that involve roleplaying child sexual abuse (see Ben Weiss, "Meta and OpenAI have spawned a wave of AI sex companions – and some of them are children", https://fortune.com/longform/meta-openai-uncensored-ai-companions-child-pornography/?utm_source=Iterable&utm_medium=email&utm_campaign=reader&tpcc=NL_Marketing).

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

Response: Yes – 5.15 and 6C.108 is confidential.

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.

Response: This is covered in our response to Question 1.

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

Response: 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?

[Is this answer confidential? Yes / No (delete as appropriate)]

We firmly support minimum requirements and basic obligations for ALL services, and are pleased to see that all services must name someone responsible for compliance with illegal content duties and reporting and compliance duties (Chapter 8), and all U2U and search services have certain requirements in Chapter 9 of Volume 3.

A key theme in our remaining responses will be the notion that the concept of basing child safety proportionally on the popularity of a product does not exist in the physical world, and should not exist online. Services should be scoped in based on risk and functionality, not size. There are several examples of very small service providers that have an incredibly outsized impact on harm. These often include file hosting services or fast-growing start up social networking applications. There needs to be flexibility in the legislation/regulations to ensure these edge cases can be scoped in as needed. Canada's recently announced legislation contemplates bringing smaller companies into scope once risk is established.

Some scaling of the exact requirements may help account for the size of the user base and the company, but governments must normalize adherence to basic online safety standards – such as keeping illegal CSAM off services – as an expected "cost of doing business" in the technology industry. In our response to question 8, we propose two fundamental priorities that should matter for all online services, regardless of size.

We support safe innovation by companies of all sizes. In the same way that we establish, for example, clear and uniform food-handling rules for food establishments whether a small kiosk or a national fast-food chain, we hope that even small services have minimum requirements to ensure they address illegal content. Basic obligations will lead to more consideration of issues during development, and help prevent online services from being misused.

Any measure of size will to some degree create arbitrary distinctions between companies that just below and just over the line. It may create confusion if a user base fluctuates rapidly, and may in some cases create incentives to disperse services over multiple platforms.

As well, services aimed at children are likely to have smaller user bases. Platforms aimed at children can attract some adults who intend to harm children, but such platforms are intended for, and generally used by, a smaller portion of the population. Such services are less likely to meet a user threshold that is based on the entire population, and a more appropriate measure would be the number of child users. This may be dealt with more in the third consultation, but for this stage, we offer the example of the Wizz app. Wizz is a chat app owned by a Paris-based company aimed, according to its website, at 13-24 year olds. The app permits users to swipe through profiles (like how some dating apps work) and chat with strangers. Cybertip.ca has

received 180+ reports concerning Wizz since 2021, leading to a Cybertip.ca alert in early 2024. As stated in the alert, compared to 2022, we received 10 times as many reports about the app in 2023. Reports about Wizz increased faster than any other platform. Most reports concerned sextortion involving male victims. The majority of victims reported to Cybertip.ca were between 15 - 17 years old. The app has 15 million users worldwide.

In the remainder of our response to this question, we will first set out the basic requirements we believe all services should have to incorporate into their business models. We will then provide a Canadian example that illustrates the importance of minimum requirements such as codes of conduct, training, and tracking complaints.

Aspects of the governance and accountability proposal on page 5 that should apply to all services:

- **Tracking new illegal content** – Any service should be obligated to track new kinds of illegal content and unusual increases in illegal content. Companies can leverage automation to do this efficiently. This relates back to our comment above about normalizing adherence to basic standards. Without tracking, there is a critical evidentiary gap surrounding illegal behaviour online and what services are being targeted.
- **Code of conduct** – A code of conduct is a basic obligation in many other contexts and will help promote safer company cultures. The code of conduct could be as simple as a one-pager that makes employees aware of the company's commitment to removing CSAM, prioritizing child safety, and having no tolerance for harassment, illegal material, etc. At this stage, awareness of safety and prevention of illegal conduct has taken a backseat to functionality, revenue generation, and other growth. Individuals focused on providing a service may not appreciate that safety and addressing illegal content are core responsibilities of operating the service and a code of conduct will help raise awareness of these considerations. We acknowledge that additional consideration of this requirement may be needed for services that are run by a single individual.
- **Adequate compliance training** – Training is essential to the success of compliance programs and should be a basic requirement on all services. The Canadian example below will illustrate the need for training and what can happen training is lacking.
- **Annual risk review – modified**: The annual risk review process is only required of large companies with more than 7 million monthly UK users. As recognized in Volume 2, smaller companies may be targeted because of their size and the perception that they have less robust safeguards. As noted in the opening of this response, smaller services aimed at children may still pose significant risks and warrant regular risk reviews. We have countless examples of smaller services being used to share CSAM and are aware that the selection of certain services can be an orchestrated action by the perpetrator community. Looking at the most common file-hosting services used to distribute CSAM based on Project Arachnid data, it is unlikely that any of these services — most of which would not be recognized by average citizens — would meet the 7 million UK user threshold. **We strongly recommend considering either a lower user threshold or a threshold based on total bandwidth. In the alternative, perhaps the annual risk review could be scaled down for smaller companies while still requiring some reporting so they turn their minds to this issue on a regular basis, as technology evolves and new risks emerge.**

Canadian example

A Canadian example of the need for policies, training, and monitoring illegal content can be seen in *R v YesUp ECommerce Solutions Inc.,* 2020 CarswellOnt 19731 (ONCJ) (only available on

Westlaw – copy provided with submission). This example also illustrates the variety of services some companies offer, and the interconnected nature of services – in this case, a website owner renting servers from another company to run a file hosting service.

YesUp Ecommerce Solutions was incorporated in 2001 by two brothers; its services included online advertising and home internet services. In 2007, the company started a file host website, and in 2011, it expanded into online web hosting. The company hired a contractor to set up required IP addresses, and the contractor listed himself as the Tech Contact and the Abuse Contact with the American Registry for Internet Numbers. **In 2012, the company drew up two abuse policies, one focused on copyright violations and one on network attacks/hacking – neither addressed CSAM or other child safety concerns.**

In 2012, Lumfile.com, a file host service run by a different individual, began renting 32 of YesUp's servers for its file hosting. Within two weeks of Lumfile.com's launch, YesUp began receiving notifications that the service was being used to host CSAM. **Between May and October 2012, YesUp received over 200 notices that their services were making CSAM available. During this time, YesUp took no action to the address the problem and in fact, continued to provide technical support to Lumfile.com.** Eventually, Canadian police became involved and seized the Lumfile.com servers.

As for the size and growth of Lumfile.com, the reported legal decision notes, "Between July and August 2012 alone, Lumfile.com saw 3,289 new members; an increase of 42%. In the same time frame, the number of files uploaded to Lumfile.com rose by 288,001, or an increase of 260%. By October 2012, Lumfile.com had 59,954 registered users, and 1,817,142 files uploaded to its servers" (paras 43 and 44).

As to the amount of CSAM on the service, over *395,994* pictures or movies of confirmed CSAM were made available for download, and over 19 million CSAM files were downloaded. Police identified 11,901 user accounts at Lumfile.com that were engaged in downloading CSAM. The numbers suggest the site was targeted by the community of perpetrators and was quickly used for illegal purposes after its launch.

YesUp was convicted in 2020 of making CSAM available and fined $100,000. The contractor who set up the servers/IP addresses and three employees were also fined $1,000. It seems that YesUp's executives were ignorant about the company's obligations to prevent the illegal activity occurring through its servers. The sentencing judge told the company, "There is a path to redemption for YesUp eCommerce Solutions Inc. If they are committed to continuing with the business model they have started, they should **embrace a more proactive approach to file hosting and set up processes and strict protocols to intervene and prevent the spread of these damaging materials**. They can learn from their error and be a leader, learning first from other leaders in the field, and changing course to improve their situation" (para 96 – emphasis added). **This illustrates the need for basic responsibilities for companies of all sizes.**

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

Response: Response covered in 3(i) above.

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

Response: No.

Question 4:	
i)	Do you agree with the types of services that we propose the governance and accountability measures should apply to?
Response: See our response to question 3.	
ii)	Please explain your answer.
Response:	
iii)	Is this response confidential? (if yes, please specify which part(s) are confidential)
Response: 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?
Response:	
ii)	Is this response confidential? (if yes, please specify which part(s) are confidential)
Response:	

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

Service's risk assessment

Question 7:	
i)	Do you agree with our proposals?
Response: As stated in response to question 3, child safety obligations cannot shift based on size. We do not tolerate that in the physical world, and we must keep that in mind when we assess proportionality, which is a consideration throughout Section 9. The following are selections from the consultation that speak to when safety should be a factor:	

- Introduction 9.7 "The adoption of good practice in risk assessment is not only a legal obligation for services, but a key component of delivering the wider industry and culture change that will put safety at the heart of services' design and decision making." **This is the preferred approach. Establishing safety as being a central part of a design will reduce the effort to add safety after the fact.**
- Referring to the Policy objectives 9.16.d) there is a general concern with the concept of proportional risk management so as to "not place undue burden on services". **When it comes to illegal material and protecting children there should be no compromise.**
- The statement in 9.31 raises concern, "Evidence from industry also supports that assessing relevant evidence is already part of some services' risk assessment practices, and that a flexible, scalable approach is most appropriate for the purposes of online safety." **This has not been the general experience of the Canadian Centre for Child Protection.** Refer to our report, Project Arachnid: Online Availability of Child Sexual Abuse Material (2021), which noted "high levels of image recidivism and the often long delays in removal times, suggest many ESPs are not deploying sufficient resources to eliminate, or at least limit, the presence of CSAM and harmful-abusive content" (page 3). We have also documented media-reported instances of companies failing meeting their own standards and/or protect their users – see "Track record of online harm".

ii)	Please provide the underlying arguments and evidence that support your views.
Response:	
iii)	Is this response confidential? (if yes, please specify which part(s) are confidential)
Response: 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?

Response: The "four-step risk assessment process" appears to capture a framework that should provide the necessary guidance for determining areas of risk and putting in place steps to monitor and mitigate risk.

Important to note is the acknowledgement in 9.62, "We are clear in the Service Risk Assessment Guidance that in some instances the number of users may be a weak indicator of risk level. They need to be considered alongside other risk factors. It is possible for a large service to be low risk, and for a small service to be high risk, depending on the specific circumstances of each service." We have observed this concern realized on the platforms OMEGLE and WIZZ. Size is not the determining factor for risk and the assessment process does acknowledge that critical point.

Although it is mentioned in other areas of the regulation, and consultation, we recommend making a clear statement in the risk assessment guidelines regarding two fundamental priorities:

- **ZERO TOLERANCE FOR ILLEGAL MATERIAL**: that there is zero tolerance for illegal material (e.g. child sexual abuse material) and a corresponding core responsibility for any service is to prevent the upload and distribution of illegal material on their service; and
- **CHILD SAFETY IS A PRIORITY**: given children are the largest most vulnerable users online, that care for their safety is a central requirement for the management of risk on any service

The risk factors that were noted and seen as important to the safety of children include:

- Table 9.4 "User complaints, including user reports": "operate a complaints procedure to a service that … (b) provides for appropriate action to be taken by the provider of the service in response to complaints of a relevant kind, and (c) is easy to access, easy to use (including by children) and transparent".
- Table 9.4 "Retrospective analysis or 'lessons learned' following incidents of harm … Services should have some kind of process in place to diagnose where and how things went wrong following any significant instances of harm." With the added recommendation that this analysis be extended to any harms not just to a subset defined as "significant".

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

Response:

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

Response: No.

Question 9:	
i)	Are the Risk Profiles sufficiently clear?
Response: The Risk Factors (that build the Risk Profiles) in the draft Service Risk Assessment Guide (Annex 5) do identify characteristics that the Canadian Centre for Child Protection would deem critical in the protection of children, including: - Clearly identifying when "illegal harm" may be a risk - Clearly identifying when harm related to CSEA may be a risk - Identifying the general risks associated with "Adult services" … NOTE: there is an added risk that could be highlighted that children, if not effectively restricted from accessing the adult service, could be harmed from exposure to the material. - Clearly identifying that services which allow child users are at risk of being accessed by the offending community. - Identifying the risks associated with anonymity.	
ii)	Please provide the underlying arguments and evidence that support your views.
Response:	
iii)	Do you think the information provided on risk factors will help you understand the risks on your service?
Response:	
iv)	Please provide the underlying arguments and evidence that support your views.
Response:	
v)	Is this response confidential? (if yes, please specify which part(s) are confidential)
Response: No.	

Record keeping and review guidance

Question 10:	
i)	Do you have any comments on our draft record keeping and review guidance?
Response:	
ii)	Please provide the underlying arguments and evidence that support your views.
Response:	
iii)	Is this response confidential? (if yes, please specify which part(s) are confidential)
Response:	

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?
Response:	

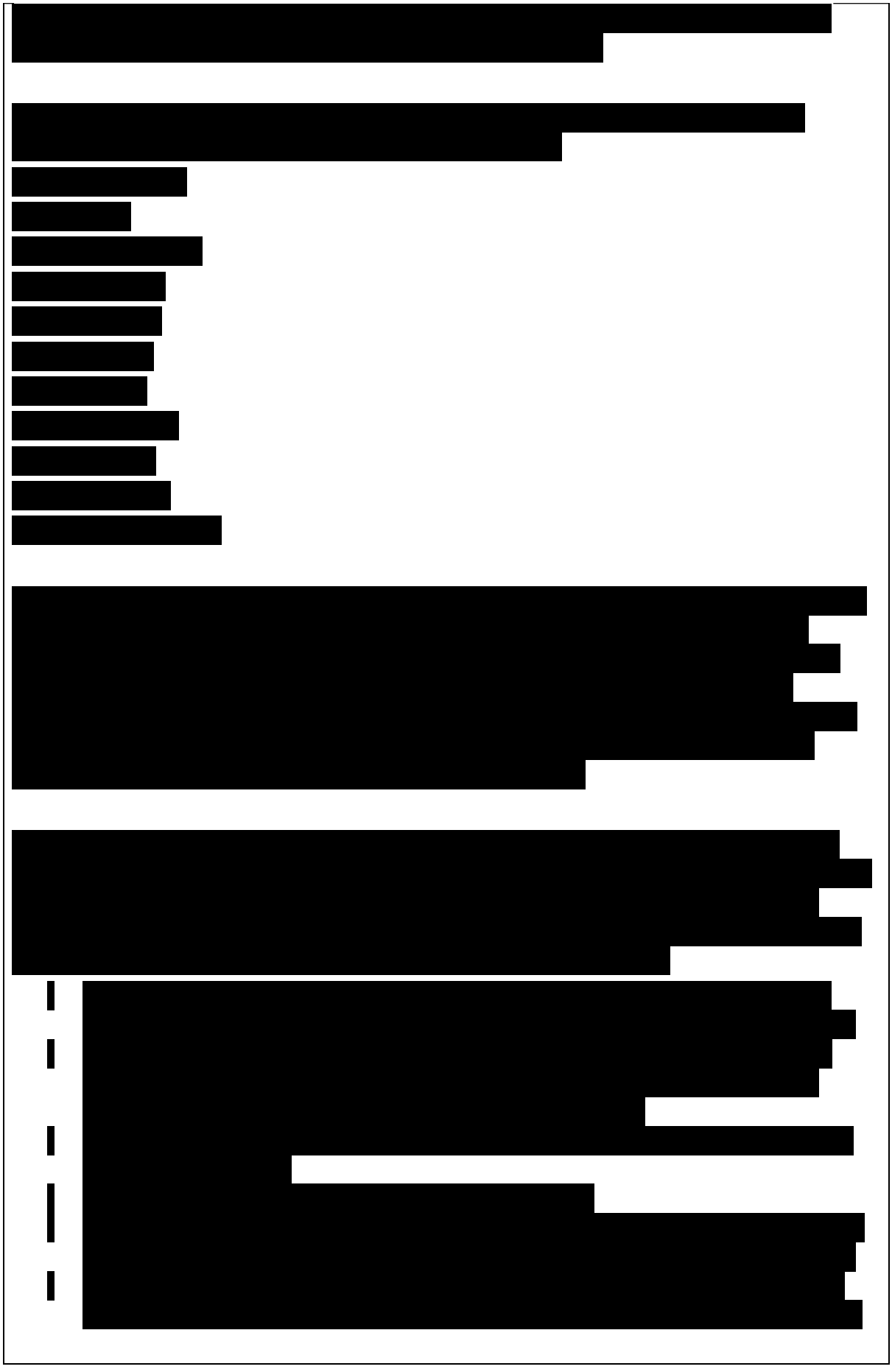
ii)	Please provide the underlying arguments and evidence that support your views.
Response:	
iii)	Is this response confidential? (if yes, please specify which part(s) are confidential)
Response:	

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?
Response: Our comments on the overall approach are similar to our comments in the risk assessment section - the concept of basing child safety proportionally on the popularity of a product does not exist in the physical world, and should not exist online. We understand the approach of placing additional requirements on larger companies as a general proposition in any regulatory framework. However, we are concerned that without guardrails and fundamental requirements for entities that allow user-generated/uploaded content, the best interests of children are being outweighed by concerns over competition and innovation. If we accept the proposition that safety stifles innovation, then we are saying unsafe companies have an advantage over companies that innovate safe products. What is needed is an approach that incentivises safe innovation across all companies rather than one that favours unsafe product design. Prioritizing child safety is consistent with the UK government's obligations under the *United Nations Convention on the Rights of the Child* to make the best interests of the child a **primary consideration** in all actions, including regulatory actions, concerning children. We note that Volume 4 indicates the proposals within are first steps. However, tech has been unregulated for far too long and the scale of harm has been too extensive to take an overly cautious approach. We have seen that industry actors have not taken voluntary action, and in the wake of this, CSAM has flourished online. Ending impunity in relation to CSAM must be an **urgent** priority.	
ii)	Is this response confidential? (if yes, please specify which part(s) are confidential)
Response: 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?
Response: No. [redacted]	
ii)	Please provide the underlying arguments and evidence that support your views.
[redacted]	



[redacted]

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

Question 14:	
i)	Do you agree with our definition of large services?
Response:	
ii)	Please provide the underlying arguments and evidence that support your views.
Response:	
iii)	Is this response confidential? (if yes, please specify which part(s) are confidential)
Response:	

Question 15:	
i)	Do you agree with our definition of multi-risk services?
Response:	
ii)	Please provide the underlying arguments and evidence that support your views.
Response:	
iii)	Is this response confidential? (if yes, please specify which part(s) are confidential)
Response:	

Question 16:	
i)	Do you have any comments on the draft Codes of Practice themselves?
Response: We have focused on our comments on Volume 4 as the Codes reflect the proposals therein.	
ii)	Is this response confidential? (if yes, please specify which part(s) are confidential)
Response:	

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

Content moderation (User to User)

Question 18:	
i)	Do you agree with our proposals?

Response: We agree that all U2U services should be required to have systems designed to take down illegal content swiftly. We recommend applying the following proposals to all U2U services (they are currently limited to large and multi-risk services):

- Prepare and apply a policy about the prioritisation of content for review.
- Ensure people working in content moderation receive training and materials that enable them to moderate content efficiently.

Having a policy and providing some training are not overly onerous requirements. The training may be less extensive for smaller companies, but there must still be some training.

Having a policy and training are both closely related to the requirement to have systems to take down illegal content, and it is difficult to envision the systems functioning effectively without a

policy and training. Volume 4 acknowledges the need link between training and giving effect to content moderation policies at 12.174. We also recommend that Ofcom provide guidance on the need for regular training. The tactics offenders use to harm children evolves, and there must be ongoing training that updates trainees on emerging risks.
ii) Please provide the underlying arguments and evidence that support your views.
Response:
iii) Is this response confidential? (if yes, please specify which part(s) are confidential)
Response: No.

Content moderation (Search)

Question 19:
i) Do you agree with our proposals?
Response: We agree that all search services should be required to have systems designed to deindex or downrank illegal content. We recommend applying the following proposals to all search services (they are currently limited to large and multi-risk services): - Prepare and apply a policy about the prioritisation of content for review. - Ensure people working in search moderation receive training and materials that enable them to moderate content effectively. Our reasoning for recommending the policy and training requirements extend to all parties is the same as in Question 18. Additionally, having a policy and training is especially important given the flexibility search services have to deindex or downrank according to their own determinations of what is appropriate in the circumstances. Missing from the proposal is a consideration of the privacy violation of the person depicted when content is not deindexed or downranked. The focus is on minimising the risk of individuals encountering illegal content through searches, but when that illegal content is CSAM or other imagery of a survivor of CSAM, there is also a need to protect that individual's privacy and safety. Protection for victims' rights is acknowledged in 12.62 in relation to removing illegal content from U2U services.
ii) Please provide the underlying arguments and evidence that support your views.
Response:
iii) Is this response confidential? (if yes, please specify which part(s) are confidential)
Response: No.

Automated content moderation (User to User)

Question 20:
i) Do you agree with our proposals?
Response: We agree that automated content moderation should be required for user to user services. However, the proposals rest on the misunderstanding that automated CSAM moderation is costly or difficult to employ for smaller companies. We will address costs under Question 22. Respecting the user thresholds on page 89, smaller services may not track their user number. For the 70,000 UK user threshold, it is very possible that some organizations are not currently set up to know this number. 70,000 UK users is also a high threshold considering the amount of CSAM that could spread on a service of that size.

As well, the notion that automated content moderation is privacy invasive overlooks the fact that allowing CSAM to be uploaded and shared is massive violation of the privacy of the person depicted.
ii) Please provide the underlying arguments and evidence that support your views.
Response: In our 2021 Project Arachnid report, we noted that nearly half of all media (48%) that triggered the issuance of a removal notification to an ESP, had previously been flagged on that ESP's service by Project Arachnid, illustrating the problem of known CSAM appearing on services that do not employ adequate moderation or upload prevention. On the need to scope in small services, we refer again to the YesUp example above. The Lumfile.com service had under 70,000 **worldwide** users, yet police identified over 11,000 accounts engaged in download CSAM; over 395,000 pictures or movies of confirmed CSAM were made available for download, and over 19 million CSAM files were downloaded. This is one example involving a Canadian provider from 2012. Imagine how much larger the numbers may be today. The file hosting service example provided in response to question 13 also illustrates on CSAM can proliferate at an immense scale from a small service.
iii) Is this response confidential? (if yes, please specify which part(s) are confidential)
Response: No

Question 21:
i) Do you have any comments on the draft guidance set out in Annex 9 regarding whether content is communicated 'publicly' or 'privately'?
Response: The delineation between public and private remains vague. We understand a pragmatic approach, but there could be some suggested thresholds or a consideration of the specific nature of groups that are formed to share CSAM. We are aware of large Telegram groups used to share CSAM. WeProtect has noted that "the use of E2EE group messaging services is likely to grow in the future… [which] could lead to wider distribution of child sexual abuse material". To the extent these groups have "access restrictions", it should be noted that those restrictions are generally intended to evade law enforcement and are in place due to the illegal nature of the group. We also reiterate that the sharing of CSAM is a violation of the privacy of depicted individuals. More than the privacy of the user/group is at stake.
ii) Is this response confidential? (if yes, please specify which part(s) are confidential)
Response: No.

Do you have any relevant evidence on:

Question 22:
i) Accuracy of perceptual hash matching and the costs of applying CSAM hash matching to smaller services;

Response:

In our experience, perceptual hash matching is highly accurate and an effective way to remove CSAM. Project Arachnid has issued over 39 million removal requests based on hash matching. False positives have been rare. In 2021, we reported that roughly half of media actioned by Project Arachnid is removed within 24 hours.

Project Arachnid handles up to 10s of millions of images every day from smaller providers. Our tool, Shield by Project Arachnid, is available to smaller companies at no cost. It is an API that companies can easily implement to help prevent CSAM from being posted and distributed on their systems. We have several smaller companies using our hash lists for free, and there is no specialist knowledge required. We understand Microsoft PhotoDNA software is also free for qualified users.

Meta Ireland's 2022 report on processing under EU Regulation 2021/1232 reveals very few false positives. See https://transparency.fb.com/sr/eu-csam-derogation-report-2023/. Meta states that it uses media matching technology and conducts human audits on samples of detected media. There were 6.6 million CSAM matches made related to EU citizens with 29,000 appeals. Of the appeals, 3,700 pieces of content were restored. This gives a real-world baseline for accuracy rates on a very high volume provider. 29,000 appeals from users equates to an appeal rate of 0.43%.

As acknowledged in 14.64, the costs are likely to be significantly lower, and perhaps even negligible for smaller services. As further acknowledged in footnote 202 of Volume 4, NGOs will often work with smaller companies to provide no or low-cost options, and that is what C3P does.

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

Question 23:	
i)	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;
Response: C3P makes its hash list available to ESPs who request access for the purpose of filtering or removing CSAM. We will work with the company to put in place the right contractual terms and other measures to enable access.	
ii)	Please provide the underlying arguments and evidence that support your views.
Response: See above.	
iii)	Is this response confidential? (if yes, please specify which part(s) are confidential)
Response: No.	

Question 24:

i)	Costs of applying our CSAM URL detection measure to smaller services, and the effectiveness of fuzzy matching for CSAM URL detection;
Response: We do not have specific evidence on the costs of CSAM URL detection or the effectiveness of fuzzy matching for CSAM URL detection. We note that Project Arachnid can match URLs, and we make our list of URLs available in a similar manner to our hash value lists, at no cost.	
ii)	Please provide the underlying arguments and evidence that support your views.
Response: N/A	
iii)	Is this response confidential? (if yes, please specify which part(s) are confidential)
Response: No	

Question 25:	
i)	Costs of applying our articles for use in frauds (standard keyword detection) measure, including for smaller services;
Response:	
ii)	Please provide the underlying arguments and evidence that support your views.
Response:	
iii)	Is this response confidential? (if yes, please specify which part(s) are confidential)
Response:	

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.
Response:	
ii)	Please provide the underlying arguments and evidence that support your views.
Response:	
iii)	Is this response confidential? (if yes, please specify which part(s) are confidential)
Response:	

Automated content moderation (Search)

Question 27:	
i)	Do you agree with our proposals?
Response: We agree that URLs which have been identified as hosting CSAM or as being part of a website entirely or predominantly dedicated to CSAM should be deindexed from the search index of a relevant service. It is important to ensure this includes cached pages. In addition, it must be recognized that problematic sites can readily shift content to new web addresses. So the initial problematic site is shut down, and within hours/days a new site emerges with the same content depicted, and so on and so on. Obligations should be in place to help ensure swift identification and deindexing of these identical sites given that this is a known phenomenon.	
ii)	Please provide the underlying arguments and evidence that support your views.
[redacted]	
iii)	Is this response confidential? (if yes, please specify which part(s) are confidential)
Response: [redacted]	

User reporting and complaints (U2U and search)

Question 28:	
i)	Do you agree with our proposals?
Response: We agree with the proposals for simple complaints mechanisms and taking appropriate action on complaints. See our comment on paragraph 16.85 below.	
ii)	Please provide the underlying arguments and evidence that support your views.
Response: In section 16.85, we agree that asking for context, while important, may engage user privacy rights. In our experience, companies have at times asked children to provide identification	

to prove their age, or in the case of non-consensually distributed intimate images, asked the victim to provide identification to prove it is them in the image/video. While companies must comply with data protection laws, not all will take the necessary care with personal information, so there must be limits to what companies can require before they take action.

In our 2020 report, "Reviewing Child Sexual Abuse Material Reporting Functions on Popular Platforms", we explained the need for CSAM-specific reporting functions within various functions of popular platforms (e.g., within a post, reporting a user, within messages/chats, and other publicly visible content when not logged in). We also noted that some survivors use these functions to report their own CSAM and flag it for removal.

[redacted]

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

Response: Yes, [redacted]

Terms of service and Publicly Available Statements

Question 29:
i) Do you agree with our proposals?
Response: The proposals are a good starting point. Terms of service are often the basis for Project Arachnid removal requests on harmful-abusive material (which encompasses harmful images of children that do not meet a criminal law threshold but may nonetheless violate an ESP's terms).
ii) Please provide the underlying arguments and evidence that support your views.
Response:
iii) Is this response confidential? (if yes, please specify which part(s) are confidential)
Response: No.

Question 30:
i) Do you have any evidence, in particular on the use of prompts, to guide further work in this area?
Response:
ii) Please provide the underlying arguments and evidence that support your views.
Response:
iii) Is this response confidential? (if yes, please specify which part(s) are confidential)
Response:

Default settings and user support for child users (U2U)

Question 31:
i) Do you agree with our proposals?
Response: We are concerned with the ability of children to change default settings. We are also concerned about the fact that these measures will only apply to those platforms that have some form of age assurance already. We recognize that an age assurance consultation will be coming next year, but in the meantime, the UK has an Age Appropriate Design Code and other tools such as the UK GDPR which ought to be leveraged to ensure meaningful change is implemented earlier than the time age assurance mechanisms are in place. Moreover, limiting these measures to platforms with a high risk for grooming and large platforms with a medium risk for grooming is too narrow. Those with medium and high risk, regardless of size, should be in scope.
ii) Please provide the underlying arguments and evidence that support your views.
Response: Considering the extent of grooming of, and harm to, children known to be occurring on platforms, awaiting the age assurance consultation before the measures in Chapter 18 are realized unduly delays the protection of children. Some of the measures proposed are not unreasonable even for

adult users so perhaps the case can be made for implementing some measures for all users until such time as reliable age assurance measures are in place.

The ability of a child to control default settings makes sense for older children age 16-17. Children under 12 should not be able to change their settings. We understand the challenges with children aged 12-15. However, our concern is the vulnerability of children in that age range to online grooming offences.

Statistics Canada, our national statistical agency, has said "**the risk of cybervictimization increases with age, from 12 to 17**, mirroring the increased frequency in the use of social networking, video and instant messaging as youth age" (https://www150.statcan.gc.ca/n1/pub/75-006-x/2023001/article/00003-eng.htm). Specific to online sexual exploitation, a Statistics Canada report released in 2022 indicated that **84%** of victims of online sexual offences against children were **aged 12-17** (https://www150.statcan.gc.ca/n1/en/pub/85-002-x/2022001/article/00008-eng.pdf?st=6-LIdPlg - see pages 8-9)

In our work, we do receive reports about the online sexual exploitation of children under 12, but the 12-14 age group is incredibly vulnerable to online luring. For sextortion, we have noted that boys aged 15-17 are the biggest target (https://www.cybertip.ca/en/online-harms/alerts/2021/sextortion-increase/).

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

Response: No.

Question 32:

i)	Are there functionalities outside of the ones listed in our proposals, that should explicitly inform users around changing default settings?

Response: The app Wizz invites users as part of the account creation process to make known their other social media handles. This can provide invaluable information to an offender. Removing this type of functionality for users under 18 is an additional protection that could be put into place.

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

Response: No.

Question 33:

i)	Are there other points within the user journey where under 18s should be informed of the risk of illegal content?

Response: Default parameters for children using livestreaming functions would also be important. For example, similar protections could be included in the livestreaming function as are set out for direct messaging. In addition, information could be displayed for the child before engaging in livestreaming of the risk, for example, that it could be recorded by the other user.

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

Response: No.

Recommender system testing (U2U)

Question 34:	
i)	Do you agree with our proposals?
Response:	
ii)	Please provide the underlying arguments and evidence that support your views.
Response:	
iii)	Is this response confidential? (if yes, please specify which part(s) are confidential)
Response:	

Question 35:	
i)	What evaluation methods might be suitable for smaller services that do not have the capacity to perform on-platform testing?
Response:	
ii)	Is this response confidential? (if yes, please specify which part(s) are confidential)
Response:	

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

Enhanced user control (U2U)

Question 37:	
i)	Do you agree with our proposals?
Response: We agree with providing users controls to block/mute users and disable comments on their own posts. However: - We are concerned that Ofcom is not recommending that services extract or retain information about blocked accounts (see section 20.44). A counter consideration is that if platforms kept track of blocked accounts, that is likely to surface bad actor accounts more quickly. Treating the complaints as one-offs misses the overall pattern.	

- We also recommend considering the interconnected nature of many platforms and the need for information to be shared across various services operated a given company.
- There should also be an opportunity to know that someone is coming from the same IP address (and therefore likely the same person) - there must be a more proactive way to protect end users in these cases.
- Finally, users should have more options for geo-fencing. They should be able to limit who can contact them outside of a certain geographical radius such as their own city.

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

Response: The reason we believe more must be done to protect users is because we see over and over again, through our operations, that youth will block the user and the user (or group of users) will simply set up a new account and recontact the youth. Perpetrators are relentless in this approach, especially in sextortion incidents. The levels of aggression and violence are startling, and it often wears children/youth down.

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

Response: No.

Question 38:

i) Do you think the first two proposed measures should include requirements for how these controls are made known to users?

Response: Yes, there should be requirements for making these controls known to users, including any changes to how the controls works or added controls. The controls themselves must be very visible and not buried.

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

Response: No

Question 39:

i) Do you think there are situations where the labelling of accounts through voluntary verification schemes has particular value or risks?

Response:

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

Response:

User access to services (U2U)

Question 40:

i) Do you agree with our proposals?

Response: In addition to the proposal included, we recommend adding "or a user connected to or part of organized crime, such as a financial or sexual extortion ring".

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

Response: Police organizations worldwide are experiencing a surge of reports tied to organized crime operating outside the reach of local law enforcement. These reports relate to extortion tactics employed against vulnerable users, particularly but not exclusively children . Many police agencies have stated publicly these crime enterprises are based in Nigeria. Regardless of where they are based, once platforms start getting reports from users about these types of tactics, in our view the platforms should be able to implement mechanisms to block these predatory users from accessing UK users.
iii) Is this response confidential? (if yes, please specify which part(s) are confidential)
Response: 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)?
Response:
ii) What are the advantages and disadvantages of the different options, including any potential impact on other users?
Response:
iii) Is this response confidential? (if yes, please specify which part(s) are confidential)
Response:

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?
Response: In C3P's view, it is reasonable for a user who has shared known CSAM to lose the privilege of ever returning to that service.
ii) Is this response confidential? (if yes, please specify which part(s) are confidential)
Response: 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?
Response: Using quality hash/URL datasets and supplementing with human moderation will assist. The combination of these two methods is particularly important when the content being classified

involves adolescent CSAM. Human moderation alone, without the additional flag associated to a hash set that indicates the youth is a part of an identified series, may otherwise result in a human moderator overlooking it. For example, Meta has publicly stated that their moderation approach is to "err on the side of an adult" if the age of the person depicted is questionable, illustrating the complete abdication of any concern for the adolescent child. See https://www.nytimes.com/2022/03/31/business/meta-child-sexual-abuse.html.
ii) Is this response confidential? (if yes, please specify which part(s) are confidential)
Response: No.

Service design and user support (Search)

Question 44:	
i)	Do you agree with our proposals?
Response: Yes with the exception that detecting CSAM searches and deploying warnings should apply to all search functions, not only large services.	
ii)	Please provide the underlying arguments and evidence that support your views.
Response: We agree with the evidence Ofcom has cited about the role search services can play in providing a pathway to CSAM. We note that section 22.40 is important. Online communities of all types develop their own vernacular and coded references; CSAM and CSEA perpetrators are no different. We add that services need a source of data that provides keywords and symbols that might be expected to return illegal results, but are associated with CSAM. This information must be updated on a regular basis as new terminology is introduced to evade detection.	
iii)	Is this response confidential? (if yes, please specify which part(s) are confidential)
Response: 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?
Response: It is important to keep in mind the survivor perspective when considering what is proportionate. For a survivor, it may not be proportionate to allow small services to harm until they become larger.	
ii)	Please provide the underlying arguments and evidence that support your views.
Response:	
iii)	Is this response confidential? (if yes, please specify which part(s) are confidential)
Response: 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?
Response: Overall, the measures are still the minimum required to address the worst kinds of harm on the services; the burden is relatively light considering the risks, and this is why we have recommended extending certain requirements (such as having policies and training) to smaller platforms.	
ii)	Please provide the underlying arguments and evidence that support your views.

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

Question 47:	
i)	We are applying more measures to large services. Do you agree that the overall burden on large services proportionate?
Response: Large services have considerable resources and capacity to do more. Considering the scale of harm that ensued when safety mechanisms were voluntary, it is not disproportionate to require significantly more transparency and precision.	
ii)	Please provide the underlying arguments and evidence that support your views.
Response:	
iii)	Is this response confidential? (if yes, please specify which part(s) are confidential)
Response: 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?
Response:	
ii)	Please provide the underlying arguments and evidence that support your views.
Response:	
iii)	Is this response confidential? (if yes, please specify which part(s) are confidential)
Response:	

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?
Response: Our response is limited to CSAM and CSEA offences. As acknowledged, CSAM offences are "are amongst the least complex priority offences analytically". This is an important message. A risk with the level of detail for some sections is that if providers believe they must assess intent to the level suggested in Volume 5, their assessments will slow down processes and may result in content remaining visible much longer than it should. Certain proposals seem to put companies in the position of assessing criminal intent and state of mind, while also signalling that companies should not dig too deep and accept content at face value because of user privacy concerns. For example, in **26.140**, there is a discussion of the potential need for information about the relationship between a perpetrator and victim and the history of their interactions. At the end of this paragraph, there is a reference to the perpetrator only being "guilty of the offence" if they know or ought to that the behaviour will have a serious effect on the victim. It should clear throughout the proposals that removal is different from a determination of criminal guilt. **26.38** – The statement that "there is no additional class of information which automated tools could have access to that human moderators could not" may be true in certain contexts. However, for CSAM, automated CSAM detection tools track and leverage previous human assessments so they can be employed at scale, reducing the need for continued individual assessment of previously identified CSAM. For example, C3P's Project Arachnid utilizes a team of trained analysts around the world to assess content and the assessments are logged with the cryptographic and perceptual hash value for the image. Once an assessment is made, there is no need for an individual at a technology company to view the material that matches that cryptographic hash (nor material that is very close in distance to a perceptual hash). The dataset that Project Arachnid uses could be considered an "additional class of information" which the system can access and which a human moderator would not have access to. **26.46** – The suggestion that content may at times be posted as a critique is not applicable for CSAM and this should be acknowledged. If a video is CSAM, it is CSAM irrespective of the intent of the poster. Within the guidance, it is important to clearly distinguish parameters for assessing CONTENT from parameters for assessing INTENT. Some criminal offences are incohoate and an analysis of intent may be required, but other types of content, such as CSAM does not transform from being CSAM to not being CSAM based on an intent analysis. **26.68** – We recommend acknowledging the role of non-illegal imagery of children, particularly material that does not meet criminal thresholds but is of an identified or known victim of CSAM (i.e. clothed or partially clothed image/video of the victim related to the abuse), in facilitating CSAM offences. In our 2019 Framework we underscored the need to remove all images in an

abusive series and noted that clothed images from the start of a CSAM video or series are used to point to where to find additional imagery involving child sexual abuse. This material may be used to "commit or facilitate the commission of an offence" and "it may nevertheless be appropriate and proportionate to remove it from services in compliance with the safety duty more generally".

26.148 – Contextual information relevant to inferring age should also include settings within the imagery (e.g. appearance of a child's bedroom in background of the image/video).

26.149 – "Hard evidence" could have different meanings for readers, including that there must be some age verification from outside of the content itself. Also, what a reasonable person assumes is someone under 18 based on appearance is very open to interpretation and context/training of those reviewing/moderating the content can influence what they believe is reasonable. Most moderators currently assess based on much stricter criteria – individuals without any sexual maturation characteristics – because of the ease of making this determination which means that material of children over 12 (the age when sexual maturation characteristics start appearing) is often left to circulate online.

Further guidance/training on this aspect will be important. The training and tools are provided to human moderators will directly affect their ability to assess/infer age.

26.150 – This section should take into account information from a reporting entity indicating the individual is under 18, and information from someone else in the child's life, who knows they are under 18 and reports to the platform. This could be a parent/guardian or even a friend. We have also seen situations where other users are aware the youth is under 18 and are trying to assist by reporting. Overall, this section relies heavily on youth reporting situations on their own and does not consider situations where a parent/guardian, friend, school counsellor, etc. may provide information to the platform. Also see also our comment on 26.153 related to age estimation tools.

26.151 – Does this also exclude the use of public posts and public profile information? Many youth are providing information that would assist in determining their age on their public profile (e.g., school, activities). We have seen situations where youth are trying to share their age in their profile but are doing it in clever ways – through the use of images with hidden numbers for their age, math problems in the textual profile information, etc. We recommend distinguishing between public information and private information that the company would have to go "under the hood" to obtain.

26.152 – There should be some parameters for what constitutes "good evidence". We have had situations where companies have found information about now-adult victims of CSAM and claimed they were over 18, even when imagery was recorded and shared when they were under 18.

26.153 – There are currently bias/gaps with age estimation tools leaving users with a false security. These tools have been shown to error in the estimation of age, creating risks for youth. It is not clear what is considered "highly effective", especially when it comes making determinations of youth who are close to age 18, and again, biases may exist that make age estimation tools less accurate based on the youth's characteristics, expression, or other factors.

26.154 – This should clearly cover obfuscation techniques for links. Examples include the use of additional characters in a URL to prevent URL detection or hotlinking (i.e.

http://www.example.☺com/) and not linking directly to CSAM through the use of link shorteners and/or redirects through multiple links.

26.159/160 – The context of discovery of the situation may assist with the determination of when the service can reasonably infer that a potential victim is under 16. For example, did someone report the activity to the platform? Was it the youth themselves, a friend, a parent/guardian, an agency/NGO/reporting entity?

26.160 – The last sentence is of concern. If the victim has made a positive statement representing themselves as 16 or over, there should still be a consideration of the overall context. Kids who are under 16 may be particularly vulnerable if they are representing themselves as over 16 or even over 18 to other users. Actual age of the child would be something to consider here, particularly with those 13 and under. This point may also be more relevant to a police investigation or prosecution to determine criminal liability; a removal decision is different. As well, in the case of applications with age requirements in their terms of service, a child underage could be considered to have represented themselves as of age simply by the fact they have circumvented the age barrier of the platform (this is recognized in 26.162).

26.165 c) – This section relies heavily on youth reporting situations on their own. What about when a parent/guardian, a friend of the youth, someone else in their life, or even an agency/NGO/reporting entity provides information to the platform about the age of the potential perpetrator?

26.166 – Similar to our comment on 26.151 above, we would advocate for a distinction between publicly and easily accessible information like profile information and back-end data that only the provider has. Again consideration of discovery situation should be taken into account, including if someone reported to company or if the company detected using their own moderation tools.

26.273-275 – suicide and self-harm – The concerns described in this paragraph also directly affect kids, especially as part of some sextortion schemes. It should be acknowledged that suicide encouragement combined with a threat to distribute an intimate image or other private information can be an indicator of illegal intent.

It is possible certain instances may be covered under other provisions re: sexual exploitation since it may go hand-in-hand with grooming but this is not always the case. However, encouraging suicide or self-harm is not always connected to sexual exploitation, or at least not directly.

Particularly in the case of a child, whose brain is considered in development until age 26, encouraging self-harm by anyone should be considered serious abuse. Children are still developing their ability to put this type of encouragement into perspective and assess it critically; they are especially susceptible to what they read. As well, consideration should be given for the age of the individual who is doing the encouraging as children are not always sufficiently mature to understand the harm they are inflicting. Overall, there may be special considerations for children in this area that should be noted here and that other civil organizations could contribute.

ii) What are the underlying arguments and evidence that inform your view?

Response: Our responses to the questions in this section are based on our experience operating Cybertip.ca, receiving thousands of reports every month of online concerns involving children from Canadians, and discussions we have had with various electronic services providers about what is illegal or harmful.

Our comments on **26.153** are informed by our experience reviewing the Wizz app after we received nearly 200 reports about it through Cybertip.ca. In an alert information the public about the app, we stated, 'Wizz's "age verification" process appears to primarily be done by applying artificial intelligence to a submitted selfie. This process is known as "age estimation", and it is far from perfect. **Female Cybertip.ca analysts who are 23 and 25 years old went through the facial recognition process and were able to create accounts on Wizz as 16-year-old males**.' In the case of Wizz, it appeared the age estimated-related safety claims by Wizz could provide a false sense of security. It is critical to understand the limitations of age estimation tools to avoid over-reliance on such tools .

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

Response: No

Question 50:

i) Do you consider the guidance to be sufficiently accessible, particularly for services with limited access to legal expertise?

Response: No. The guidance is overly legal, confusing, and appears to require assessment of criminal intent.

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

Response:

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

Response: No.

Question 51:

i) What do you think of our assessment of what information is reasonably available and relevant to illegal content judgements?

Response:

26.26(b) – Reference to companies having information from "trusted flaggers". In our experience, only a few of the big tech companies have "trusted flaggers" and it is based on their invite/criteria to determine who is trusted. In the context of CSAM and harmful material of children, we recommend considering language like 'reporting entity or body' in addition to trusted flagger or clarifying the definition of trusted flagger so it is not based on company discretion.

26.26(d) - This should also include information about networks/connections and comments/posts by others connected to the user. Context here is often very valuable for assessing intent.

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

Response: 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?
Response:	
ii)	Please provide the underlying arguments and evidence that support your views.
Response:	
iii)	Is this response confidential? (if yes, please specify which part(s) are confidential)
Response:	

Enforcement powers

Question 53:	
i)	Do you have any comments on our draft Online Safety Enforcement Guidance?
Response:	
ii)	Please provide the underlying arguments and evidence that support your views.
Response:	
iii)	Is this response confidential? (if yes, please specify which part(s) are confidential)
Response:	

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