Consultation response form

Please complete this form in full and return to protectingchildren@ofcom.org.uk.

Consultation title	Consultation: Protecting children from harms online
Organisation name	UK Safer Internet Centre

Overview of the UKSIC

The UK Safer Internet Centre (UKSIC), established in 2011, is a leading global partnership helping to make the internet a great and safe place for everyone. We provide support and services to children and young people, adults facing online harms, and professionals working with children.

A bridge between Government, industry, law enforcement and society, we are the engine of the online protection landscape in the UK, dealing with both prevention and response.

Formed of three charities, Childnet, Internet Watch Foundation and SWGfL, we work together to identify threats and harms online and then create and deliver critical advice, resources, education and interventions that help keep children and young people, and adults, safe. We share our best practices across the UK and globally.

We focus our work around four functions:

- **An awareness centre:**

Where we provide advice and support to children and young people, parents and carers, schools, and the children's workforce.

- **Three helplines:**

Which provide support to professionals working with children and young people with online safety issues, and support to all adults facing issues with harmful content and non-consensual intimate imagery online.

- **A hotline:**

Which provides an anonymous and safe place to report and remove online child sexual abuse images and videos wherever they are found in the world.

- **A voice to young people:**

We operate a Youth Advisory Board, and we nurture youth participation, providing a focus on youth voice to give young people agency to make a difference in their school communities.

UKSIC is the proud coordinator of Safer Internet Day in the UK.

Our partners the IWF and the SWGfL have submitted responses to this consultation and we fully endorse their responses. We have summarised and integrated their responses where appropriate. Please refer to their full responses for the detail.

Summary

As the UKSIC we would first like to congratulate the efforts of Ofcom in undertaking such a complicated and comprehensive task. We would also like to acknowledge the efforts

of Ofcom to make the consultation response process more accessible and comprehensible.

We also want to acknowledge the positive steps that have been taken. We are pleased to see Ofcom making progress and appreciate the high level of engagement with children. The combination of age assurance measures and new measures relating to recommender systems represents a significant advancement in protecting children.

Furthermore, compared to the illegal harms code, the children's consultation is easier to navigate and more concise.

Our response aligns with our submission to Ofcom's illegal harms consultation, as our analysis and feedback remain consistent.

Approach to risk and Governance

In general, we are pleased with the attempt to limit the online harms that children face, and in particular the introduction of age assurance measures for all services.

At the same time, by applying the proposals from the illegal harms' consultation to the children's consultation which includes a focus on size over risk, we are concerned with the potential implications on smaller services.

Though we comprehend the approach and classification of risk by size, UKSIC believes that smaller platforms can also pose several risks including: Intimate Image Abuse, Harassment, CSAM hosting and others which will be covered throughout the response. A safety-by-design principal approach should ensure that smaller and larger platforms are designed to be safe for the users, while also ensuring that they comply with any regulations.

To ensure safety by design, regulation should focus on small but high-risk platforms, and Ofcom's risk assessment approach should encompass not only large platforms, where best practices often exist, but also medium-sized companies.

The current approach, which limits safety measures to companies with a medium to high risk of CSAM, is inadequate. A review of the definition of "large platforms" is necessary to include some of the most popular platforms used by children and to ensure that medium-sized businesses are also included in training and development requirements for staff. Ofcom must enforce protections on small but high-risk platforms to ensure robust safety measures across all services.

Safety-by-design issues

Throughout this consultation we will provide data, evidence and examples on ways that the safety-by-design framework could be included to mitigate risk and minimise the harm that people face online

The current focus and strategy of this consultation is placed upon the notion of minimising the burden of the industry. However, from our work with the Helplines and Hotline and our awareness centre we have first had experience and evidence on the harm that people can face online. We therefore propose, the inclusion of a safety-by-design framework that will cover small/large new and existing services.

Recommender systems also have a significant effect on the content that children consume in online platforms, and we are not entirely satisfied with the provisions in the RS category. More specifically, RS1 which aims to "Ensure that content likely to be Primary Priority Content is not recommended to children" we propose to prevent Primary Priority Content to being recommended to any user, completely removing csam, intimate image abuse and terrorist content from platforms. With the current provision and age assurance technologies which have not yet been deployed, children may still easily falsify their age, allowing them to access inappropriate and harmful content. RS2 also refers to the reduction of priority content rather than the removal of such content within the recommender feed system. Children will still be able to view content that is harmful, and could have intense psychological impact on them including, suicide content, body dysmorphia, and online challenges which have cost the lives of so many children in the UK in the past.

We therefore urge Ofcom to ensure that content likely to be Priority Content is reduced in prominence on children's recommender feeds

Age Assurance

There is no doubt that the combination of age assurance measures and new measures relating to recommender systems are significant steps forward in increasing protections for children, particularly in reducing their exposure to - and the impact of - Primary Priority Content and Priority Content that is harmful and, in some cases, life-threatening. However, the limitations of these measures in addressing broader safety by design factors persist, compounded by the safe-harbour compliance threshold which does not prioritise overall improvements in child protection.

For example:The age gating requirement is an additional obligation and the only substantial new measure to protect children, making it a single point of failure. The risk assessment obligations in this consultation are no stricter than those proposed in the illegal harms consultation, nor do they require any significant redesign of services based on identified risks. This means that services can fulfil their obligations simply by keeping children off their platforms, thereby diluting the Act's mandate to "design and operate" safer services to ensure a "higher standard of protection for children than for adults."

ADR and offering children the right to recourse

The lack of a structured approach to alternative dispute resolution (ADR) in the proposals represents a missed opportunity to bolster user trust and platform accountability significantly. ADR offers numerous advantages, such as easing the burden on formal complaints processes, fostering more positive relationships between platforms and users,

and potentially resolving conflicts in a manner that respects the interests of all involved parties. Additionally, ADR mechanisms like mediation, arbitration, or ombudsman services can bring expertise and impartiality that may not always be present in platform-driven complaints procedures.

SWGfL suggests that the proposals could be enhanced by explicitly integrating ADR mechanisms into platforms' strategies for addressing complaints and disputes. An outline of an ADR solution previously proposed by SWGfL in the illegal harms consultation; this could be supplemented by the development of specific guidance or standards for ADR mechanisms within the context of online harms. This would include criteria for mediators or arbitrators and processes that ensure fairness, transparency, and accessibility.

Referencing Report Harmful Content, the Draft Online Safety Bill (Joint Committee), in December 2021[1] recommended (paragraph 457) that; "The role of the Online Safety Ombudsman should be created to consider complaints about actions by higher risk service providers where either moderation or failure to address risks leads to significant, demonstrable harm (including to freedom of expression) and recourse to other routes of redress have not resulted in a resolution" and that "We suggest that the Department look to Report Harmful Content as a potential model for what such an Ombudsman could look like".

While the proposals in Chapter 16 establish a framework for reporting and complaints, the integration of ADR mechanisms could significantly improve the effectiveness, accessibility, and user trust in these processes. Leveraging SWGfL's expertise in online safety underscores the pivotal role ADR can play in the broader ecosystem of reducing online harm and resolving disputes.

Media literacy and education

We notice a lack of emphasis on media literacy and education for children. We want to ensure that children are effectively educated where they are online and that the platforms are held to account on educating children in an appropriate way to use their platform safely.

Whilst the platforms themselves are not well positioned to provide totally impartial education they can support children on making best use of their platforms.

As UKSIC we strongly believe in the importance of media literacy as a framework that will help online users and especially the most vulnerable remain safe online. UKSIC brings a collaborative and multifaceted approach to online safety we have greatly understood the importance of media Literacy. More specifically, through Childnet's Digital Leader's Programme and the Youth Advisory Board we work closely with children to understand their

[1] SWGfL, 2021 https://swgfl.org.uk/magazine/report-harmful-content-release-final-quarterly-report-for-2021/

unique needs in terms of support and media literacy. Additionally, SWGfL one of the partner organisations of UKSIC, operates project evolve which is a digital education toolkit that covers knowledge, skills, behaviours and attitudes across eight strands of our online lives from early years right through to eighteen. These outcomes or competencies are mapped to age and progress. The statements guide educators to the areas that they should be discussing with children as they develop their use of online technology.

We also believe that there needs to be a strong read across from formal and non-formal education in terms of children's online safety education and that platforms should seek to work with NGO's and other to reinforce key messages.

Your response

Question	Your response
Volume 2: Identifying the services children are using Children's Access Assessments (Section 4).	
Do you agree with our proposals in relation to children's access assessments, in particular the aspects below. Please provide evidence to support your view. 1. Our proposal that service providers should only conclude that children are not normally able to access a service where they are using highly effective age assurance? 2. Our proposed approach to the child user condition, including our proposed interpretation of "significant number of users who are children" and the factors that service providers consider in assessing whether the child user condition is met? 3. Our proposed approach to the process for children's access assessments?	Confidential? – N **Proposed Approach** We would first like to acknowledge the efforts of the Ofcom online safety team in taking further steps on corporate accountability. In particular the focus on governance, accountability and management is a very important step forward which should reflect on the existing measures in other industries including the automobile and aviation safety procedures. At the same time, we would like to share our thoughts and concerns regarding the limitations of the current proposed provisions, reflecting also on our response for the illegal harms' consultation. The proposals for governance oversight are retrospective – reviewing the process of risk management retrospectively (what the company is going to do to mitigate the risks as they arise) rather than engaging in prospective analysis, looking at results from a risk assessment of the design and safety of their service and the risks of harm that may arise from it and putting mitigating measures upfront.

Question	Your response
	We would like to see online safety outcomes front and centre of accountability structures to ensure that not only are T&S staff accountable for profits but also accountable for the safety of users and they are measured accordingly. The BEEF survey[2] highlights the importance of measuring user experiences relating to safety and holding T&S and senior staff accountable. UKSIC shares the concern of SWGfL who mention in their response:*"SWGfL do not believe that internal monitoring is sufficiently independent. Platforms should be monitored by an external independent auditor to maintain independence Page 5 and impartiality and therefore public trust in the maintenance of platforms as safe spaces."* UKSIC therefore proposes the introduction of an external independent auditor similar to the ICO investigation period[3] to maintain independence and impartiality. We are concerned that the initial assessment t relies on the self-assessment as we know that self-assessment has failed for many years to see companies deploy effective child safety measures. As the UKSIC we are supporting the proposal put forward by the Online Safety Act Network which entails that Ofcom should shift away from prescriptive "tick-box" approach and instead adopt a duty of care principle for the children who are using online platforms. Reflecting on the illegal harms response which is directly linked with the "Protecting Children from Illegal Harms" consultation, we are still concerned with the overt importance of size of the service in terms of the safety provisions proposed by Ofcom. We have provided further evidence on

[2]Beef, Complaint Ex. 1_To Be Sealed_MT-IG-AG-NM-000220597 (courtlistener.com)

[3]ICO https://ico.org.uk/about-the-ico/our-information/our-service-standards/#:~:text=We%20aim%20to%20resolve%2090,line%20with%20the%20customer%20charter.

Question	Your response
	"Developing the Children's Safety Codes: Our framework (Section 14)". **Helplines and Trusted Flagger** As the official UK Intimate Image Abuse helpline (Revenge Porn Helpline) we have first-hand experience with victims of online abuse, who have sought out help through our service. We would therefore like to see the introduction of a Trusted Flagger system, where the helplines operated by SWGfL and other legitimate stakeholders could share information, practices and be in constant communication with Ofcom and online services. This process could also act as an extra online safeguarding step which will reinforce or refute the findings of the service provider self-assessment and child user condition document. We would also like to point Ofcom to the direction of the Guidelines and best practice for the trusted flagger relationship between helplines and online service providers document which was developed by the Early Warning Working Group of the UK Council for Internet Safety, and published in April 2023, see https://www.gov.uk/government/publications/trusted-flagger-programmes Further down in a response we are analysing our current working partnership with Tik Tok which swiftly takes action taking down harmful content. **Super Complaints** There are significant resource implications to the body that is eligible to make a super-complaint, at the same time the resource and expertise implications can be significant for organisations, to the point that it could prevent them from putting forward any submissions due to complexity and costs. An organisation could take a significant time to be able to build conclusive evidence and build a file report that matches the description that is provided by this draft.

Question	Your response
	While the proposals in Chapter 16 of the Illegal Harms Consultation establish a framework for reporting and complaints, the integration of ADR mechanisms could significantly improve the effectiveness, accessibility, and user trust in these processes. Leveraging SWGfL's expertise in online safety underscores the pivotal role ADR can play in the broader ecosystem of reducing online harm and resolving disputes. Additionally, we think that Ofcom should streamline the process to ease the entry of organizations who are planning to submit a sufficient and full super-complaint. Accordingly, the Super Complaints entity requirement outlined in S.169 of the Online Safety Act, sets out Ofcom as the sole responsible body to make the inspection and provide the final verdict. That fact leads to the lack of an alternative body for any individual complaints, while also intrinsically delaying the process and outcome leading to aggravated harm to the victims.
Volume 3: The causes and impacts of online harm to children **Draft Children's Register of Risk (Section 7)**	
Proposed approach: 4. Do you have any views on Ofcom's assessment of the causes and impacts of online harms? Please provide evidence to support your answer. a. Do you think we have missed anything important in our analysis? 5. Do you have any views about our interpretation of the links between risk factors and different kinds of content harmful to children? Please provide evidence to support your answer.	Confidential? – N Overall, we are pleased to see the definitions of harmful content and that children will not only be protected from illegal content but content that is harmful to them. We do stress though that this links to effective age assurance and effective safer by design principles that need to be widely adopted. We do urge Ofcom to keep a watching brief on emerging issues and risks and to respond dynamically especially where out helplines see increased reports relating to harms to children. <u>Pornography</u>

Question	Your response
6. Do you have any views on the age groups we recommended for assessing risk by age? Please provide evidence to support your answer. 7. Do you have any views on our interpretation of non-designated content or our approach to identifying non-designated content? Please provide evidence to support your answer. **Evidence gathering for future work:** 8. Do you have any evidence relating to kinds of content that increase the risk of harm from Primary Priority, Priority or Non-designated Content, when viewed in combination (to be considered as part of cumulative harm)? 9. Have you identified risks to children from GenAI content or applications on U2U or Search services? a) Please Provide any information about any risks identified 10. Do you have any specific evidence relevant to our assessment of body image content and depressive content as kinds of non-designated content? Specifically, we are interested in: a) (i) specific examples of body image or depressive content linked to significant harms to children, b. (ii) evidence distinguishing body image or depressive content from existing categories of priority or primary priority content. 11. Do you propose any other category of content that could meet the definition of NDC under the Act at this	We agree that pornography should be included as a Primary Priority Content. As UKSIC we are concerned with the widespread availability of pornographic content and the potential harmful effect it can have on children. Research tells us that accidental exposure is the main reason young people come across this content. Volume 3 recognises that "The average age at which children first encounter pornography is 13, although older children (14-17) are more likely to see it regularly.". At the same time we would also like to reinforce a notion that is included throughout volume 3, that children below the ages of 13 are present in the platforms as they have bypassed the existing ineffective age assurance methos. In fact, The BBFC Revealing Reality Report[4] provides examples on how children as young as 7 years old came in contact accidentally with pornography. "The majority of the young people interviewed in the qualitative research had seen pornography by the age of 13. There were examples, however, where respondents reported viewing pornography as young as 7 years old. One interviewee, Chanelle (18), saw pornography via a pop-up on an illegal streaming website when she was 7, while April (18) says she used to regularly search for and watch "violent" 5 pornography at 7 or 8 years old, after initially stumbling across it accidentally. A very small number first viewed pornography much later, at 17 or 18." Children are able to access pornography, and age-inappropriate content without the foundation media and sexual health literacy that will allow them to separate fact from fiction. For many Young People and children pornography acts as their source of Sex Education, which has major potential implications in terms of understanding what healthy sexual relations are, expectations about what

[4] https://www.revealingreality.co.uk/wp-content/uploads/2020/01/BBFC-Young-people-and-pornography-Final-report-2401.pdf

Question	Your response
stage? Please provide evidence to support your answer.	bodies look like and what sex is like, as well as issues that are existing in the industry including: violence, sexism and racism. With regards to sex education, we would also like to reinforce the role of the school system in partnership with parents. Since the availability of such harmful content has already been so widespread the need for an effective and transparent RSE system is of paramount importance in order the bridge the gap between fiction and reality. Research for Safer Internet Day 2023 found that children aged 8-17 are more likely to talk to a parent or carer above anyone else (e.g. a teacher) if they have any concerns about something they see or experience online. However, that same research also found that over a third of parents and carers (36% in total) are unsure of where to turn for support if faced with a challenging situation about their child's online life. Guidance which encourages schools to build a positive relationship with parents will mean that key educational messages can be reinforced at home and in school and will have the added benefit of helping empower parents to access support and guidance when they need it. Additionally, through the Professional Online Safety Helpline operated by SWGfL, we have witnessed over 700 cases of Harmful Sexual Behavior Support requests in schools, and therefore as UKSIC we would reinforce the call for the introduction of school policies that address Harmful Sexual Behavior in schools. <u>Safety By Design</u> Age appropriate experiences can only be achieved with safety-by-design principles that will ensure

Question	Your response
	that children who are legally present on the platforms are not facing harm from the content that is easily accessible. Age assurance measures should therefore also be used on existing underage users of the platforms in order to minimise the harm and risk whilst they use the service. <u>End-to End Encryption</u> We are pleased to note the recognition in Volume 3 (7.11.55) that End-to-End Encryption (E2EE) is identified as a feature carrying specific risks, particularly concerning its facilitation of perpetrators disseminating child sexual abuse material while minimizing the risk of detection. This assertion is strongly supported by robust evidence base derived from police-recorded crime statistics[5] , the firsthand experiences of victims of such crimes, and the legal proceedings involving prolific offenders like David Wilson. Had Facebook Messenger employed End-to-End Encryption, it is highly probable that Wilson would have eluded detection, thereby leaving the 500 boys he communicated with and the 51 boys he coerced into sharing indecent images of themselves potentially unsafeguarded. At the same time, we are sharing a concern regarding the wording of End-to-End encryption that we also shared in our Illegal Harms Consultation response. UKSIC therefore urges Ofcom to carefully consider the implications of classifying End-to-End Encrypted services as private communications providers. Such a classification could lead to unforeseen long-term consequences, potentially prompting social media networking sites to shift their encrypted services into the "private" category to either evade their obligations under the Act or circumvent the expenses associated with content moderation.

[5] https://www.nspcc.org.uk/about-us/news-opinion/2024/Child-abuse-image-crimes-increase-calling-ofcom-tech-companies-take-action/

Question	Your response
	Revenue Models and Safety-by-Design UKSIC recognises the inclusion of business models and revenue models as a significant risk and functionality that could harm children and supports volume 3 p (7.12.5). However, we are quite concerned with the overall omission of Safety-by-Design principles and the advice on services to adopt such processes and business models. Instead, we are proposing a shift in the focus from the industry-centric approach of this consultation to a victim-centred and child-centred approach which would enhance the provisions that are set out. For instance, social media appears to have significant negative effects on the mental health of children who are users. In the 2024 Safer Internet Day research[6], there is evidence that children are affected negatively when using such platforms: (36%) of children notice a negative change in their mental wellbeing when they are online. Notably, the proportion of young people who sometimes notice this negative change is highest among both younger children and older teens, with 38% of 9-to 10-year-olds on average and 39% of 15- to 17-year-olds on average feeling this way. These figures for younger children are striking given that the minimum user age requirement for the social media platforms they are mostly using is 13. Effective age assurance measures and safety by design functionality choices which will age differentiate the content accessible by different age brackets is a necessary step to mitigate the harm caused by addictive-design recommender functionality choices. Recommendation

[6] Research - UK Safer Internet Centre

Question	Your response
	As mentioned in our previous responses, we advocate for a stronger emphasis on encouraging services to develop platforms that are safer by design. At present, the Codes are overly focused on content, necessitating a shift towards consistent, proactive measures to prevent and disrupt harm. As Ofcom notes in Volume 2 (5.20), services prioritising growth often neglect safety measures, leaving them susceptible to exploitation by CSEA perpetrators. Therefore, the Codes should place more emphasis on early-stage interventions, such as employing proactive technologies to detect illegal and harmful content and implementing measures targeting perpetrator behaviour. A pragmatic, precautionary regulatory approach is essential, focusing on long-term safety by design. This approach should apply equally to all user-to-user services, including those offering end-to-end encrypted communications. Since private messaging is a primary channel for online grooming, it is crucial to design these communications to be safe for children. However, as the proposed measures only apply to public spaces, we are concerned that perpetrators will shift their activities to private spaces. Without applying safety by design to all user-to-user communications, this approach does not mitigate the risk but merely relocates it. Steps must be taken to ensure children are protected in private and end-to-end encrypted environments. In addition to our concerns about end-to-end encrypted environments and private messaging, Volume 3 highlights issues with audio and live streaming, which are not adequately addressed despite being identified as harmful functionalities in the children's safety duty risk register. Where a

Question	Your response
	risk cannot be sufficiently mitigated, services should prevent all children, or children in certain age groups, from accessing the relevant features or functionalities.
Draft Guidance on Content Harmful to Children (Section 8)	
12. Do you agree with our proposed approach, including the level of specificity of examples given and the proposal to include contextual information for services to consider? 13. Do you have further evidence that can support the guidance provided on different kinds of content harmful to children? 14. For each of the harms discussed, are there additional categories of content that Ofcom a) should consider to be harmful or b) consider not to be harmful or c) where our current proposals should be reconsidered?	Confidential? – N <u>Age Assurance and Self-Assessment</u> One point of concern we have identified relates to measure AA3 and AA4, which state that services "Whose principal purpose is not the hosting or the dissemination of one or more kinds of Priority Content, and which do not prohibit one or more kinds of Priority Content" will have to employ effective age assurance measures. If we take the example of Youtube for instance which principal purpose is not the hosting of priority content but also prohibit violent content and nudity would that mean that they don't have to utilise age assurance measures according to their self-assessment? Service providers are also not 100 per cent accurate in blocking or taking down primary priority or priority content but may claim to do so in the self-assessment, leads to a gap in the self-assessment process and the actual user experience in the platform. An example of our work can be seen in a recent case involving YouTube and the Professional Online Safety Helpline. It was reported to POSH that there was a significant amount of nude and sexually explicit content appearing on YouTube when users typed a series of 'x' into the search function. We alerted YouTube, who asked us to send specific links to the content. Initially, we reported over 30 pieces of content, most of which were removed. However, hundreds of similar pieces of content remained active under the search term. We advised YouTube of this issue,

Question	Your response
	but they requested that we sift through the content and make individual reports. Due to the enormous volume of videos with such violations, we decided that we did not have the resources to continue reporting, as it would have taken days. Moreover, once one video was removed, another was uploaded. We had alerted Google to the trend, and they were aware of the search terms used to find the content but resisted investigating further. Unfortunately, the content remains available on the platform to this day. The need for an independent and alternative checks and balance system is of paramount importance and actors/services who reduce harm online should play a significant role in the design and evaluation process of platforms functionality and safety measures. <u>Harmful Online Content</u> As UKSIC we are concerned with the online content that is easily accessible by children on social media. We would also like to acknowledge the inclusion of recommender systems alongside advertising business models as a function that can greatly exacerbate the risk of accessing online harmful content. Children are affected by the content they consume online, so measures need to be taken to ensure that their experience online is safe and age appropriate. In recent years, particularly with the introduction of short video form content, the effect of function systems plays a significant role in children. In a recent study published by UCL[7] there was clear evidence that hateful ideologies and misogynistic tropes that were shared online and massively spread with the help of the algorithm, have moved off screens and into schools, becoming embedded in mainstream youth cultures. Vodafone[8] also

[7] https://www.ucl.ac.uk/news/2024/feb/social-media-algorithms-amplify-misogynistic-content-teens

[8] https://www.vodafone.co.uk/newscentre/press-release/ai-aggro-rithms/

Question	Your response
	conducted research which showcased significance evidence that AI recommender systems are probing young people into harmful and extremist content; “on average, boys aged 11-14 are exposed to harmful content within 30 minutes of being online and one-in-10 are seeing it in as little as 60 seconds. This worrying trend stems from AI algorithms pushing content promoting misogyny (69%) or violence (79%) to boys following innocent and unrelated searches (59%)”. The Safer Internet Day research[9] provides also insight into the understanding of children online of recommender systems: 71% of children that participated in the SID 2024 research told us: “we understand that when they ‘like’ or watch something online, it influences what content is suggested to them in future. 62% understand that algorithms choose the content they see in their feed or games and videos that are recommended to them. Particularly worrying also for extreme pornography which is easily accessible by children and often present in social media platforms which are largely used by children. Particularly alarming issue is the ease of access to harmful content on social media platforms popular with children. This poses significant risks related to several priority offences, including threats, abuse, harassment, and the potential escalation to terrorism offences. Recommender systems can indoctrinate children into extremist views by continuously pushing such content. Indeed, reports such as those by Ribeiro et al. (2020)[10], and Amnesty International (2023) have highlighted how the affordances of social media platforms, such as YouTube's recommender systems, actively amplify and direct harmful content. This includes content associated with the alt-right, misogynist and manosphere content (Reset

[9] https://d1xsi6mgo67kia.cloudfront.net/uploads/2024/02/UK-Safer-Internet-Day-2024-Research-Report.pdf
[10] Ribeiro et al. 2020, The Evolution of the Manosphere across the Web | Proceedings of the International AAAI Conference on Web and Social Media

Question	Your response
	Australia, 2022)[11], and self-harm material (Amnesty International, 2023)[12]. In a worldwide first, a recent ruling by a UK coroner in October 2022 identified Instagram as 'likely contributing' to a young person's death due to the high level of self-harm material recommended on her feed in the months before she died. Online echo chambers and the rapid rise and circulation of electoral fake news, mis/disinformation, and polarisation towards political extremism on the internet have had particular impacts on young people (Lewis-Kraus, 2022)[13]. Most recently, misogynist influencers, who have leveraged their fame to promote polarised far-right extremism, have utilised the phenomenon of echo chambers to bring gendered hate to prominence in other dimensions of public discourse, such as boys' misogynist behaviours within schools and educational settings (Das, 2022)[14]. It is therefore of great importance to ensure that current social media platforms particularly large ones, provide a safety-by-design framework for the operation of their recommend systems which are safe for children, while at the same time providing the technological foundation for smaller organisations to use to ensure that in turn their recommend systems do not harm children. Overall, recommender systems which are addictive-by-design harm children, and safety should be the primary focus which is implemented by design and through proactive measures to minimize the harm caused. <u>Functionalities and inequalities</u>

[11]Reset Australia, 2022 EMBARGOED REPORT Algorithms as a weapon against women (reset.tech)
[12] Amnesty International, 2023 Driven into Darkness: How TikTok's 'For You' Feed Encourages Self-Harm and Suicidal Ideation - Amnesty International
[13] Lewis-Kraus, 2022 How Harmful Is Social Media? | The New Yorker
[14] Das, 2022 https://www.theguardian.com/technology/2022/aug/06/revealed-how-tiktok-bombards-young-men-with-misogynistic-videos-andrew-tate

Question	Your response
	Online platforms which are global, reflect existing global economic inequalities and can pose significant risks on romance scams, sextortion or other forms of harm. The cultural context is really significant, and it is something we consider greatly at the Revenge Porn Helpline.: Whilst volume 3, recognises the additional contexts which can exacerbate the harm and impact caused, there is little detail of marginalised groups and culturally sensitive content. The severity of consequences of intimate image abuse within diverse cultural groups is vital to understand, the risks of honour-based abuse, honour killings and community ostracization should be considered. The case study delves into the qualitative exploration of the profound impact that both Intimate Image Abuse (IIA) and online harms can have on a client coming from a culturally sensitive background. Our client found herself in a distressing situation when her intimate images were maliciously shared online by an ex-partner. The Revenge Porn Helpline successfully removed 3067 of these images, and an additional 188 impersonation accounts spanning Facebook, X, Instagram, TikTok, and YouTube were reported for removal by Report Harmful Content" Additionally, the 'Digital Misogynoir Report: Ending the dehumanising of Black women on social media'[15], showcases that minority and ethnic minority groups are facing multifaceted risks while online. Women and particularly Black Women are a lot more likely to be abused, harassed online, and to receive hate comments. It is therefore evident that stronger accountability should be requested by tech companies to tackle and mitigate for the rise of hate comments and abusive rhetoric

[15] https://glitchcharity.co.uk/research/#:~:text=The%20Ripple%20Effect%20Report%201%2C800%2C000%20people%20suffered%20threatening,this%20has%20sadly%20increased%20during%20the%20Covid-19%20pandemic.

Question	Your response
	that affects minorities online. As evident in volume 3 ethnic minorities and women appear to face disproportionate harms online. Should these be taken into account for the risk profiles (geographical distribution of the users). Platforms with users with extreme socio-economic inequalities without proper provisions could provide a fertile ground for grooming and sextortion[16]. The approach to proportionality As the UKSIC we would like to reinforce the issues raised in the Online Safety Act Network response[17]. **Issue: Proportionality Focus on Economic Considerations:** Ofcom's approach to proportionality appears predominantly economic, aiming to avoid imposing costs on companies. While the Online Safety Act (OSA) mandates that regulated services adopt a proportionate approach in fulfilling their duties, considering provider size and capacity, it also requires attention to levels of risk and the nature and severity of harm. Proportionality should balance the economic impact on companies with the societal costs and prevalence of harms to users, including impacts on the criminal justice system and support services for victims, particularly for women, girls, and minority groups. **Severity of Harm Consideration:** The severity of harm involves not only the number of affected individuals but also the intensity of the impact. Despite recognizing harms in the risk register, Ofcom's code of practice measures do not explicitly consider these aspects. The current fo-

[16] https://www.weprotect.org/issue/livestreaming/

[17] OSA Network, 2024 20240716 - OSA NETWORK CHILDREN'S CONSULTATION RESPONSE (onlinesafetyact.net)

Question	Your response
	cus is more on the economic burden on tech companies rather than balancing it against the societal costs of harm. **Small vs Large Companies: Misjudged Proportionality:** The proportionality analysis assumes that smaller companies pose less harm due to their limited reach. However, this assumption overlooks the severe harm that can occur to minoritized groups on targeted small sites. The Act includes 53 references to "proportionate," emphasizing that measures should be proportionate to the risk of harm rather than merely considering company size or capacity. **Parliamentary Debate Insights:** During the Lords Committee stage debate, the Government Minister assured that the child safety duties would be tailored to the size and capacity of providers. Smaller providers still need to meet child safety duties if their services pose a risk to children. These providers must implement systems and processes reflecting their services' risk level, ensuring they achieve the required child safety outcomes. **Ofcom's Proposal Extracts: Cost vs Benefit Analysis:** Ofcom's consultation documents indicate that impacts on services, including costs, are important to ensure requirements are justified. The documents mention that high-cost burdens might reduce investment in user safety or drive services to stop operating in the UK, impacting both children and adults. This cost consideration should not overshadow the need to keep children safe. **Recommendation:**

Question	Your response
	UKSIC recommends that Ofcom should ensure a balanced approach in its proportionality analysis. Measures should equally weigh the severity of harm and the societal costs associated with online safety issues. Furthermore, proportionality should not solely focus on economic considerations but should encompass a comprehensive understanding of the potential harms and their broader impacts on society. UKSIC is also concerned with emerging technologies and the potential risks that could impose on Children. Most notably A.I the risks will also increase exponentially. A new report[18] published by the IWF illustrates that A.I poses a significant risk particularly with the potentially exacerbated volume of CSAM mages that will require a thorough and comprehensive process to remove such content. Nudifying and deepfake technologies are also particularly worrying, including the scope of the illegal harms consultation as most of the generative A.I technologies and service providers would be considered as" small" due to their user size. UKSIC would therefore agree with the call of global cooperation that IWF proposed in 2023[19], that should reflect a global online safety regime, where the risk and harm will be minimised. Additionally, there is a notable gap in the absence of any measures in the codes related to livestreaming, especially since the risk register identifies this functionality as causing harm in several areas covered by the children's safety duty. Additionally, back in 2021, the DCMS specifically included practical guidance for companies on livestreaming in its "Principles of Safer Online Platform Design."

[18] IWF, 2024 How AI is being abused to create child sexual abuse material (CSAM) online (iwf.org.uk)
[19] UKSIC 2023, Global collaboration needed as thousands of AI-generated child sexual abuse images emerge depicting the worst kinds of abuse - UK Safer Internet Centre

Question	Your response
	By addressing these concerns, UKSIC believes that Ofcom can implement a more effective and balanced regulatory framework that aligns with the objectives of the Online Safety Act and ensures the protection of all users, particularly children, across all service providers, irrespective of their size.
Volume 4: How should services assess the risk of online harms? **Governance and Accountability (Section 11)**	
15. Do you agree with the proposed governance measures to be included in the Children's Safety Codes? a) Please confirm which proposed measure your views relate to and explain your views and provide any arguments and supporting evidence. b) If you responded to our Illegal Harms Consultation and this is relevant to your response here, please signpost to the relevant parts of your prior response. 16. Do you agree with our assumption that the proposed governance measures for Children's Safety Codes could be implemented through the same process as the equivalent draft Illegal Content Codes?	Confidential? – N We broadly agree with the proposed governance arrangements set out in the proposal. We think it's crucial that effective governance and accountability measures are in place to protect children. The current provisions assume that companies will adopt a goodwill and transparent self-assessment process which unfortunately is not reflecting reality. Arturo Bejar, the Meta whistleblower who has recently testified to the US Congress, observed: "Social media companies are not going to start addressing the harm they enable for teenagers on their own. They need to be compelled by regulators and policy makers to be transparent about these harms and what they are doing to address them."[20] The use of a trusted flagger system, where organisations who operate on the ground including ours with the Revenge Porn Helpline and Report Harmful Content, can provide a transparent independent assessment system that will provide an extra safety and mitigation measure beyond the self-assessment provided by the services. In addition, relating to point GA6 we would reinforce the need for effective training and protection measures for the moderators who are working

[20] OSA Network, 2024 20240716 - OSA NETWORK CHILDREN'S CONSULTATION RESPONSE (onlinesafetyact.net)

Question	Your response
	with children. We think that the current provisions are not sufficient particularly as many moderators and trust and safety teams have direct access to children therefore, they ought to be subjected to the checks that offline safeguarding professionals are subjected to including enhanced DBS checks and other forms of protection and safety measures. Furthermore, we believe there should be independent mechanisms in place to monitor the effectiveness and integrity of these systems (GA4). Relying solely on internal oversight is inadequate. An independent monitoring framework would provide a more objective and comprehensive assessment of the safeguards and practices in place, ensuring they meet the highest standards of child protection. This approach not only enhances transparency but also instils greater confidence in the measures being implemented to protect children online. **Trusted Flagger system** We at the UK Safer Internet Centre (UKSIC) and South West Grid for Learning Trust (SWGfL), operate three helplines (POSH, RHC, RPH), and act as trusted flaggers for several online platforms. This role enables us to escalate content for review with moderation teams, providing additional context as to why the content breaches their guidelines. Typically, our clients have already reported the content, but the report has failed to achieve the desired outcome; If we concur that the content breaches the guidelines, we then escalate it. In exceptional cases, we may report content without requiring our clients to do so first. Our relationships with these platforms have been built over years, requiring an extensive understanding of each platform's guidelines. To maintain efficiency and trust, we ensure that we only escalate content that genuinely breaches site

Question	Your response
	guidelines. The IWF operates an analyst team which identifies and reports CSAM content to platforms and other authorities. This approach requires an established system of trust and a relationship of transparency and collaboration. These relationships also involve acting as critical friends; we provide feedback on new features or policy decisions and share intelligence on emerging trends and issues. This collaborative approach helps inform their work and improves the overall safety and experience for all users. We believe that Ofcom proposing and organising a similar structure is an effective way to mitigate the gap between the self-assessment findings and the actual reality in the platforms. Impartial organisations and services can act as safeguarding actors who will identify gaps in the moderation, reporting and taking down process, eventually reducing the volume of inappropriate content in the platforms that is accessible by children.
Children's Risk Assessment Guidance and Children's Risk Profiles' (Section 12)	
17. What do you think about our proposals in relation to the Children's Risk Assessment Guidance? a) Please provide underlying arguments and evidence of efficacy or risks that support your view. 18. What do you think about our proposals in relation to the Children's Risk Profiles for Content Harmful to Children? a) Please provide underlying arguments and evidence of efficacy or risks that support your view.	Confidential? – N In general, we agree with the thorough analysis conducted by the Ofcom team and the link between functionalities and risk. However, we are still concerned with the self-assessment of corporations whose commercial interests are founded on shareholders which do not operate with a safety principle at heart. As the risk assessment is linked to the children's access assessment, we need to ensure that it is robust According to the POSH report in 2023[21], most enquiries relate to either online reputation (46%) or cyberbullying (36%). Online platforms due to the "virality" nature of their design and quick spread

[21] SWGfL, 2024 POSH Report 2023 (swgfl.org.uk)

Question	Your response
Specifically, we welcome evidence from regulated services on the following: 19. Do you think the four-step risk assessment process and the Children's Risk Profiles are useful models to help services understand the risks that their services pose to children and comply with their child risk assessment obligations under the Act? 20. Are there any specific aspects of the children's risk assessment duties that you consider need additional guidance beyond what we have proposed in our draft? 21. Are the Children's Risk Profiles sufficiently clear and do you think the information provided on risk factors will help you understand the risks on your service? a) If you have comments or input related to the links between different kinds of content harmful to children and risk factors, please refer to Volume 3: Causes and Impacts of Harms to Children Online which includes the draft Children's Register of Risks.	of information can act as an amplifier and enabler of cyberbullying in schools. Children from minority groups or SEN children would be at an increased risk to be targeted, amplifying the existing inequalities and harmful behaviours. Effective structures of content removal and redress are therefore necessary to protect the most vulnerable users. We would also like to acknowledge the inclusion of messaging services (1b) and anonymity (2b) in the risk factor analysis and in particular encryption services as a design choice which makes detection and reporting more complex, and as recognised in the risk register may be more likely to be used by children sharing violent content. We therefore call services to refrain from adopting end to end encryption messaging services. **Sextortion** Scamming and sextortion should be included in primary content category (PC) as both of these areas present significant risks to children. We have recently held a multi stakeholder round table on sextortion and shared some of the key data. In our work with the Revenge Porn Helpline, sextortion prevails as one of the most significant issues. Out of the cases the Revenge Porn Helpline supported, sextortion remained the predominant form of intimate image abuse reported to the Helpline, totalling 34% of the reports received in 2023. Overall, cases of sextortion were up 54% when compared to 2022, and cases continued to disproportionately affect men, making up 93% of sextortion reports.
Volume 5 – What should services do to mitigate the risk of online harms **Our proposals for the Children's Safety Codes (Section 13)**	

Question	Your response
Proposed measures 22. Do you agree with our proposed package of measures for the first Children's Safety Codes? a) If not, please explain why. **Evidence gathering for future work.** 23. Do you currently employ measures or have additional evidence in the areas we have set out for future consideration? a) If so, please provide evidence of the impact, effectiveness and cost of such measures, including any results from trialling or testing of measures. 24. Are there other areas in which we should consider potential future measures for the Children's Safety Codes? a) If so, please explain why and provide supporting evidence.	Confidential? – N **Safe Harbour Provision** In general, we are very supportive of the inclusion of Children's Safety Codes, and the enforcement approach that Ofcom is taking. Nevertheless, we would like to reiterate the concerns raised by IWF regarding the design of the codes as a "safe harbour." The rules-based nature of these codes may lead services to abandon existing protective or mitigating measures, under the assumption that they are no longer necessary for compliance. This could disincentivise good practices and fail to improve safety and protection standards. We recommend that Ofcom includes a requirement in the Code of Practice for all services within scope to address harms identified in their risk assessments related to features and functionalities, using best practices, even if Ofcom has not yet established an evidence base to support these recommendations. Implementing additional measures not identified by Ofcom could, for example, enhance the detection of previously unidentified CSAM content. This approach would encourage innovation in response to identified risks and support the regulation of emerging technologies like Generative AI and Extended Reality, helping to future-proof the regulation. We also support the view from the OSA Network as quoted below: *There is also a welcome warning to services - contained in volume 4 on risk assessment - that if they are "already implementing measures such that they assess their risk level to be low or negligible, they should continue doing so. Stopping implementing such measures or changing them may*

Question	Your response
	constitute a significant change (see Step 4 below) and may increase their risk level." (volume 4 pp56-57). This (to an extent) addresses concerns raised in response to the first consultation that the tick-box, prescriptive onlinesafetyact.net - 7 approach to measures in the codes - aligned with the safe harbour promise - could mean services making a decision to stop using existing protective or mitigating measures as they were no longer required to be compliant with the regulation. This is a very welcome shift. However, in terms of upholding age terms and conditions, the proposal is to measure this on a tick-box consistency metric rather than outcomes. *Again, the rules-based nature of the Codes - specifying specific recommended measures rather than obligations aimed towards the achievement of desired outcomes - and the fact that these are designed as a "safe harbour" (eg if companies follow the measures they will be judged to have complied with their duties under the Act*), means that there is no incentive for companies to implement mitigating measures beyond those described in the codes. This is the case even if their risk assessment has flagged that their service poses particular risks from other functionalities (arising from design choices) and despite the fact that the risk assessment notes the need for voluntary actions over and above what is set out in the codes. The Atlantic Council makes this point: "if compliance replaces problem-solving, it establishes a ceiling for harm reduction, rather than a floor founded in user and societal protection." (p 36)* *(*The "safe harbour" provision is described here: onlinesafetyact.net - 43 "Services that choose to implement the measures we recommend in Ofcom's Children's Safety Codes will be treated as complying with the relevant children's safety as well as their reporting and complaints duties. This*

Question	Your response
	means that Ofcom will not take enforcement action against them for breach of that duty if those measures have been implemented. This is sometimes described as a "safe harbour. However, the Act does not require that service providers adopt the measures set out in the Children's Safety Codes, and service providers may choose to comply with their duties in an alternative way that is proportionate to their circumstances ." (Para 13.4))

Developing the Children's Safety Codes: Our framework (Section 14)	
25. Do you agree with our approach to developing the proposed measures for the Children's Safety Codes? a) If not, please explain why. 26. Do you agree with our approach and proposed changes to the draft Illegal Content Codes to further protect children and accommodate for potential synergies in how systems and processes manage both content harmful to children and illegal content? a) Please explain your views. 27. Do you agree that most measures should apply to services that are either large services or smaller services that present a medium or high level of risk to children? 28. Do you agree with our definition of 'large' and with how we apply this in our recommendations? 29. Do you agree with our definition of 'multi-risk' and with how we apply this in our recommendations? 30. Do you agree with the proposed measures that we recommend for all services, even those that are small and low-risk?	*Is this answer confidential? No* **Size and risk** A key issue that UKSIC has identified exists in the classification and division of large and small services. The internet can be a particularly dangerous place for Children and the current provisions which identify large services as those with 7 million users, feel does not create a regime and framework that will effectively protect children who are using platforms and services that are considered "small". Notably, Roblox and Fortnite[4] would be excluded, which have millions of children users. As 5rights suggested, UKSIC also proposes the revision of the size criteria to 2 million monthly users to guarantee that more platforms are included within the scope of the risk mitigation. As Lord Minister Parkinson of Whitley Bay said: "I want to be clear that a small platform that is a font of illegal content cannot use the excuse of its size as an excuse for not dealing with it" [5]. Safety and innovation can co-exist, and the regulation and processes must keep their users safe and most importantly vulnerable groups such as children. By establishing a system that exempts numerous services from extensive responsibilities, Ofcom risks regressing in online safety efforts. The notion that small services are inherently safe is flawed, and companies with 7 million users should not be considered just large. We contend with the proposal of 5Rights[28] that any company with over 2 million UK users should qualify as large. The current risk classification omits several large profile companies such as Roblox and Fortnite where the user size is quite young and therefore vulnerable to risks and harms.

	Moreover, we advocate for additional clarification regarding the frequency with which services should assess their user base to identify when they've reached large-scale status. It's essential to ensure that they promptly implement additional measures for compliance once they meet the criteria. This again brings us to the question of the external auditor and how the lack of one could result into an ineffective audit and monitoring process.
Age assurance measures (Section 15)	
31. Do you agree with our proposal to recommend the use of highly effective age assurance to support Measures AA1-6? Please provide any information or evidence to support your views. a) Are there any cases in which HEAA may not be appropriate and proportionate? b) In this case, are there alternative approaches to age assurance which would be better suited? 32. Do you agree with the scope of the services captured by AA1-6? 33. Do you have any information or evidence on different ways that services could use highly effective age assurance to meet the outcome that children are prevented from encountering identified PPC, or protected from encountering identified PC under Measures AA3 and AA4, respectively? 34. Do you have any comments on our assessment of the implications of the proposed Measures AA1-6 on children, adults or services?	Confidential? – N We commend the use of Age Assurance measures, and we think it is a crucial step in providing an age-appropriate user journey. We want to see platforms adopting the most effective form of age assurance, which is inclusive and accessible. It needs to be tested in terms of bias, effectiveness. Existing users should also be checked especially those under 18. **Safety-By-Design** As the UKSIC we would like to reiterate what was mentioned in the Online Safety Act Network Response[22] to this consultation. With the exception of the proposals around recommender systems, which are welcome, the topics and measures discussed here do not significantly extend beyond the ex-post measures outlined in Ofcom's illegal harms consultation. In fact, two-thirds of the 36 measures recommended for user-to-user (U2U) platforms, and all but one of the 24 measures for search services, are either identical or equivalent. Age assurance measures, such as keeping children off platforms, are tools to prevent harm but do not constitute a "safety by design" approach that fundamentally alters the platform for all users, including children. We direct Ofcom to the

[22] OSA Network, 2024 20240716 - OSA NETWORK CHILDREN'S CONSULTATION RESPONSE (onlinesafetyact.net)

a) Please provide any supporting information or evidence in support of your views. 35. Do you have any information or evidence on other ways that services could consider different age groups when using age assurance to protect children in age groups judged to be at risk of harm from encountering PC?	analysis by the 5 Rights Foundation/Children's Coalition regarding age assurance proposals. Content moderation addresses content that has already been posted, rather than tackling the underlying system that enables its dissemination. In the "Proposed Codes at a Glance" section, the description of measures emphasizes their limitation to restricting children's access to the service (through age assurance) for Primary Priority Content (PPC) and some Priority Content, followed by more granular content-level access restrictions using age assurance, and then age verification to assess recommender system usage, alongside content moderation. This approach applies safety tech on top of an inherently harmful system rather than redesigning it for safety, especially for the users that the regulatory framework aims to protect, at a higher standard than for adults. We discuss the age assurance measures in more detail in section nine. **Age Verification and Age Assurance** We recognise the limitations of the Online Safety Act, which differentiates between adults and children by defining a child as anyone under 18. Given this definition, the current code imposes a blanket age restriction for those under 18 and does not require services to provide age-appropriate experiences for different age groups within this range. We are concerned that this one-size-fits-all approach does not consider the varying needs of children at different stages of development. We urge that services should be required to specify a minimum age requirement in their terms and conditions and enforce it effectively and ensure that children are able to understand them. The Act emphasises the 'consistent' application of age verification rather than its effectiveness. This

	focus on consistency means that even when services identify significant risks to children, the Codes do not require them to effectively mitigate these risks. The current guidance allows services to document their actions without demonstrating the actual outcomes or changes resulting from those actions. Code of Practice measures must be outcomes-based, addressing all identified risks of harm to children. Age-appropriate access to content, features, and functionalities should be established, beyond merely protecting children from 18+ content. To address this, we suggest that Ofcom align and expand the definition of a child with the Age-Appropriate Design Code. The GDPR and DPA 2018 specify that if you rely on consent for any aspects of your online service, you need to obtain parental authorisation for children under 13. Since 13 is the age of digital consent, it is crucial to prevent children of this age from accessing inappropriate services and to ensure default privacy settings protect them from grooming. As highlighted in our response to the Illegal Harms Codes, effective age assurance measures are essential for strengthening grooming mitigations. Safety-by-design measures for children's accounts are ineffective if they rely on self-declared ages that can be easily circumvented. AA3 is open to interpretation as platforms such as Meta who prohibit pornographic content, there are still reports of existing PPC content on the platforms. Would that mean they would have to implement age assurance. How will Ofcom respond if the self-assessment claims they don't allow PPC content, but helplines, hotlines and other stakeholders provide evidence that this type of content exists. The current provisions which include id provisions, passport and banking maybe excluding children who do not possess a passport or a bank

	account. Is age assurance appropriate for every user or only those that self-declare as under 18?
Content moderation U2U (Section 16)	
36. Do you agree with our proposals? Please provide the underlying arguments and evidence that support your views. 37. Do you agree with the proposed addition of Measure 4G to the Illegal Content Codes? a) Please provide any arguments and supporting evidence.	Confidential? – N **Content Moderation** <u>CSAM</u> UKSIC also acknowledges as it is mentioned in the CSAM content will not be considered as a "viral" priority content for review by social media companies who utilise automated content moderation tools. And therefore, since CSAM and grooming are both considered a priority offence, this should also reflect in the upcoming moderation processes that social media companies establish. By creating good practice which combines an automated and manual content moderation with an effective process which includes hash/matching, URL matching and a cross industry keyword list, could all contribute to a more effective content regulation. In February 2024, a study conducted by Joel Scanlon from the University of Tasmania assessed the effectiveness of the reThink chatbot project. This initiative, a collaboration between the Internet Watch Foundation, the Lucy Faithful Foundation, and Aylo (the parent company of Pornhub), has been operational on the Pornhub website in the UK since March 2022, with data collection continuing until September 2023. The reThink chatbot builds upon previously successful deterrence messaging campaigns implemented on the site since March 2021, aiming to direct potential offenders to seek assistance from the Lucy Faithful Foundation.

	During the evaluation period, key findings revealed that 99.8% of sessions did not trigger the chatbot. However, the chatbot was still displayed a staggering 2.8 million times between March 2022 and August 2023. This led to 1,656 requests for more information from the Stop It Now services, 490 click-throughs to the Stop It Now website, and approximately 68 calls to the anonymous counselling service. Before the chatbot's launch, warning messages about potential offending behaviour were displayed over 2 million times, with over 4.4 million triggers during the evaluation period. The report highlights several successful outcomes, including a significant statistical decrease in searches for Child Sexual Abuse Material (CSAM) on Pornhub UK. Additionally, most sessions that triggered the chatbot did so only once, and sessions that initially began with a search for CSAM content subsequently engaged with the site but searched for content less frequently than other sessions. We are dismayed by Ofcom's decision not to recommend any measures specifically aimed at detecting previously unidentified child sexual abuse material. We also share the IWF concerns that the current regulatory proposals set a low regulatory standard for the initial draft of the code of practice, especially considering that many companies falling under the regulation's scope already employ classifier technology to detect such material and grooming approaches. We find it unacceptable for this crucial measure to be deferred to future iterations of the Codes of Practice due to purported lack of evidence, especially when it is already considered best practice within the industry. <u>Safety By Design</u>

	The principal of safety by design in content moderation is of paramount importance to the UKSIC. We are really disappointed with Content Moderation measures (n (Volume 5, Section 16) which mostly applies solely to large and multi risk services, excluding platforms which are used by millions of children. We are also concerned specifically with point GA7 "Ensure staff involved in the design and operational management of service are sufficiently trained in approach to compliance with children's safety duties" which only applies for Large/Multi-risk Services. The omission of smaller services in the proposed provision entails that staff that may directly influence the online experience and compliance of children online will not be sufficiently trained. We are therefore calling for the expansion of the proposed measure to include all service providers that are accessed by children.
Search moderation (Section 17)	
38. Do you agree with our proposals? Please provide the underlying arguments and evidence that support your views. 39. Are there additional steps that services take to protect children from the harms set out in the Act? a) If so, how effective are they? 40. Regarding Measure SM2, do you agree that it is proportionate to preclude users believed to be a child from turning the safe search settings off? The use of Generative AI (GenAI), see Introduction to Volume 5, to facilitate search is an emerging development, which may include where search services have integrated GenAI into their functionalities, as well as where standalone GenAI services perform search functions. There is currently	Confidential? – N

limited evidence on how the use of GenAI in search services may affect the implementation of the safety measures as set out in this code. We welcome further evidence from stakeholders on the following questions and please provider arguments and evidence to support your views: 41. Do you consider that it is technically feasible to apply the proposed code measures in respect of GenAI functionalities which are likely to perform or be integrated into search functions? 42. What additional search moderation measures might be applicable where GenAI performs or is integrated into search functions?	

User reporting and complaints (Section 18)	
43. Do you agree with the proposed user reporting measures to be included in the draft Children's Safety Codes? a) Please confirm which proposed measure your views relate to and explain your views and provide any arguments and supporting evidence. b) If you responded to our Illegal Harms Consultation and this is relevant to your response here, please signpost to the relevant parts of your prior response. 44. Do you agree with our proposals to apply each of Measures UR2 (e) and UR3 (b) to all services likely to be accessed by children for all types of complaints?	Confidential? – N <u>Blocking</u> Taking into account the safety-by-design principle that was forementioned, the ability to block users should be a default setting to reduce the risk of harassment, bullying, or even the contact routes with children that could lead to grooming or self-generated intimate images. Another key point that stood out was the provision of block functionality to users of large services that identify medium or high risk. The blocking tool is a crucial tool for young people. something that in Childnet's research. Children said in our own research[23] that it is more likely to be used than reporting. Given that it is immediately effective in a way that reporting isn't, perhaps this is not surprising. We would therefore recommend that

[23] Childnet 2021, https://www.childnet.com/blog/young-peoples-views-on-reporting-online-harms/

a) Please confirm which proposed measure your views relate to and explain your views and provide any arguments and supporting evidence. b) If you responded to our Illegal Harms Consultation and this is relevant to your response here, please signpost to the relevant parts of your prior response. 45. Do you agree with the inclusion of the proposed changes to Measures UR2 and UR3 in the Illegal Content Codes (Measures 5B and 5C)? a) Please provide any arguments and supporting evidence.	blocking tools should be a requirement for all services including small services. Childnet's research in 2021 showcased that "Many young people find blocking is a useful tool in response to being worried or upset about something online – they are more than twice as likely to block someone online (44%) as report them (21%). Only 17% of 11-year-olds said they would report". This clearly showcases the importance of ensuring that blocking remains an option for children in all platforms including small services"[24]. Blocking is a more effective tool, and the omission of it in the recommended functions, provides a" fertile" and dangerous ground of grooming, harassment, cyberflashing which could all harm children significantly. Another key point that stood out was the provision of We propose the following 2 additions to the codes: • **Child safety reporting:** A significant portion of the reporting and complaints process is now automated, lacking sufficient access to human intervention. This makes it challenging for individuals, particularly parents of children, who are concerned about the impact of content on vulnerable individuals, to urgently raise such concerns. A reporting system should swiftly connect users to a human representative when a child is involved, and subsequently take necessary measures to ensure their safety. Automated systems often overlook the context in which content is displayed and to whom, thus impeding contextual judgments. Additionally, for non-registered users, services should be obligated to provide clear guidance on how to report without requiring an account setup. • **Right of appeal:** While guidelines specify how services must offer appeals to users or

[24] Childhub, 2021 childhub.org/sites/default/files/attachments/Reporting Research Final.pdf

	concerned parties who may have had content unfairly removed, it fails to include recommendations for users to appeal decisions not to remove content. Ofcom should suggest that services provide a mechanism to appeal such decisions, especially when they involve harm or risk to a child.

Terms of service and publicly available statements (Section 19)	
46. Do you agree with the proposed Terms of Service / Publicly Available Statements measures to be included in the Children's Safety Codes? a) Please confirm which proposed measures your views relate to and provide any arguments and supporting evidence. b) If you responded to our illegal harms consultation and this is relevant to your response here, please signpost to the relevant parts of your prior response. 47. Can you identify any further characteristics that may improve the clarity and accessibility of terms and statements for children? 48. Do you agree with the proposed addition of Measure 6AA to the Illegal Many children are unaware of the full implications of the terms and conditions they agree to when signing up for popular social media platforms like Facebook, YouTube, Snapchat, WhatsApp, and Instagram. To address this, simplified terms and conditions guides will be distributed to thousands of teachers across England, equipping them with the tools to educate students on understanding their online rights and responsibilities. These guides are specifically designed to empower children by providing them with clearer, more accessible information, enabling them to make informed decisions about their digital interactions. In response to growing concerns, the Commissioner is calling on social media giants to increase their transparency and accountability, particularly in how they present their	Confidential? – N **Terms of Service** Terms of Service as mentioned in the supporting documents should be accessible and should ease the user journey whilst on the platform. A lot of the services who are accessible by children as young as 13 years old, include complicate sign up and reporting processes which are incomprehensible by children who are legally using the platform. To reflect on the proposals put forward by the Children's Commissioner we request that Ofcom follows the guidelines provided which include • Awareness of the full implications of the terms and conditions they agree to when signing up for popular social media platforms like Facebook, YouTube, Snapchat, WhatsApp, and Instagram. • Simplification of the terms and conditions for all users with underage users, to enhance accessibility • Creation of Guides which will be distributed to thousands of teachers across England, equipping them with the tools to educate students on understanding their online rights and responsibilities. These guides are specifically designed to empower children by providing them with clearer, more accessible information, enabling them to make informed decisions about their digital interactions. • Increases transparency and accountability, particularly in how they present their terms of service to younger users, ensuring that these platforms are safer and more comprehensible for children.

terms of service to younger users, ensuring that these platforms are safer and more comprehensible for children. Content Codes? a) Please provide any arguments and supporting evidence.	
Recommender systems (Section 20)	
49. Do you agree with the proposed recommender systems measures to be included in the Children's Safety Codes? a) Please confirm which proposed measure your views relate to and provide any arguments and supporting evidence. b) If you responded to our illegal harms consultation and this is relevant to your response here, please signpost to the relevant parts of your prior response. 50. Are there any intervention points in the design of recommender systems that we have not considered here that could effectively prevent children from being recommended primary priority content and protect children from encountering priority and non-designated content? 51. Is there any evidence that suggests recommender systems are a risk factor associated with bullying? If so, please provide this in response to Measures RS2 and RS3 proposed in this chapter. 52. We plan to include in our RS2 and RS3, that services limit the prominence of content that we are proposing to be classified as non-designated content (NDC), namely depressive content and body image content. This is subject to our consultation on the	*Is this answer confidential? / No* **Safety-by-design** Ofcom's risk register suggests that for the majority of illegal activities covered by the legislation – such as grooming, incitement to suicide, harassment, stalking, threats, and abuse – are not amplified by the business model itself and therefore the nature of a service is not considered a significant risk factor. Instead, various features like recommender systems are identified as potential risks. However, there is substantial evidence indicating that features designed to retain user attention are inherently linked to the business model. By exempting business models from scrutiny, there's effectively a legitimization of commercial practices that are known to pose risks and cause harm, which contradicts the original intent of the legislation. As articulated by Lord Minister Parkinson of Whitley Bay: "Obligations on services extend to the design and operation of the service. These obligations ensure that the consideration of risks associated with the business model of a service is a fundamental aspect of the Bill."[7] Additionally, the proposed measures that address the recommender system come quite late in the product development and design process. A more robust "safety by design" approach, combined with rigorous risk assessment and product safety testing, should consider many more aspects of the overall service much earlier in the process. <u>Primary Priority Content and Violent Content</u>

classification of these content categories as NDC. Do you agree with this proposal? Please provide the underlying arguments and evidence of the relevance of this content to Measures RS2 and RS3. • Please provide the underlying arguments and evidence of the relevance of this content to Measures RS2 and RS3.	We would like to see violent content to be included in the Primary Priority Content. RS2 is particularly worrying, as children will still be able to view violent content within recommender systems since the provisions aims to reduce the content rather than block it or remove it. The current trends are worrying, and significant provisions should be introduced the minimise the sharing of violent and harmful content. In the past year, Report Harmful Content[25] has seen a significant rise in animal abuse reports across social media channels. Report Harmful Content can reveal that 36% of the content escalated to industry partners between November to December 2023 included animal abuse, with the vast majority involving the abuse of monkeys. Since then, Report Harmful Content has successfully been able to remove 84% of this harmful content being shared across social media platforms. The content, which is often distressing to watch, usually involves monkeys being physically and psychologically hurt and mistreated. Concerningly, Report Harmful Content has found that alongside much of this content, many viewers have actively engaged in and encouraged the torture of monkeys, revealing a concerning trend across global social media platforms. According to the POSH 2023 report, during March 2023, there was a lot of attention and a volume of cases around "TikTok riots" in schools[2], which reflects concerns across the education sector about the use of social media platforms to host content related to behaviour in schools. These incidents coincide with a number of cases specifically referring to these incidents, with enquirers requesting help to get content taken down. Regarding the location of content that caused concern (and was often requested to be removed), it

[25]SWGfL, 2024 Report Harmful Content Sees Concerning Rise in Animal Abuse Content (swgfl.org.uk)

	is no surprise, that the overwhelming majority was hosted on social media platforms. We therefore would like the reinforce the need to blocking and removing access to violent content which can have significant psychological and mental effects on children who come in contact. **Sextortion** Platforms should prevent the creation of fake profiles, and it is far more common to be targeted by gangs who are using fake profiles. The functionalities platforms allow gangs to exploit them and target young boys on popular platforms. For instance, certain functionalities including the lack of the profile verification system permits the existence of sextortion cases which target vulnerable users online. The UKSIC has hosted an Insight Research Series[26] on the topic of sexual exploitation (sextortion), and the consensus was that significant steps should be taken to protect children, and it requires a collaborative approach that will bring together, the Police, Government, NGOs and other stakeholders. The IWF hotline published, its findings which indicate a significant rise in the cases of sexual exploitation of children: "in the first six months of 2023 reports of confirmed child sexual abuse involving 'sextortion' surged by 257%* compared with the whole of 2022"[27]. **Business Model** The business model is discussed in the risk assessment and risk profiles, receiving more attention than in the previous consultation. However, the codes of practice do not address measures to mitigate or limit the commercial incentives for creating content, such as clickbait farms or harmful

[26] UKSIC 2023, https://saferinternet.org.uk/blog/insight-series-highlights-global-online-challenges
[27] IWF, 2024 https://www.iwf.org.uk/news-media/news/hotline-reports-shocking-rise-in-the-sextortion-of-boys/

	influencers like Andrew Tate, where content generation is driven by profit and often represents the creators' primary reason for being on the platform. In the risk register, Ofcom specifically references the recent rise of Andrew Tate, highlighting the financial incentives to create and share harmful content. It notes how the monetisation incentive, combined with recommender systems, leads to harmful content being pushed to younger users without their prior engagement. In the Illegal Harms Consultation, we raised our concern regarding the lack of a transparent profile verification process could have significant effects on the type of content that children consume online. The rise of AI technologies could greatly exacerbate the volume of information and fake news, whereas an effective and credible verification system could significantly assist in the distinction between misinformation and reality. According to the BBC Bitesize research[28], 37% of young people would trust influencers online as a primary source of information, and the verification system could take advantage of the trust children place on the verification scheme. If a service implements a profile verification service and a paid-for-verification service, we propose improved public transparency for users about what verified status means in practice. Children's developing cognitive abilities mean that they may struggle to discern between reliable and unreliable information online. According to Ofcom's findings, verification schemes can be exploited by malicious actors to impersonate official sources and deceive users. Specifically, reporting on X Verification has revealed vulnerabilities to scams within these schemes. Ofcom's research[29]indicates that nearly a quarter (23%) of

[28] BBC Bitesize, 2023, Young people believe influencers more than politicians when it comes to news - BBC Bitesize

[29] Ofcom, 2023 Children and parents: media use and attitudes report 2023 - Ofcom

	children express confidence in their ability to distinguish between real and fake online content, yet they struggle to identify fake social media profiles when presented with them. Given this susceptibility to fraud and malicious actors, Ofcom should ensure that services take this into account in their operations. Furthermore, any measures implemented by services to enhance transparency regarding how users can obtain verified status must be age appropriate. They should be designed to ensure that the information provided is understandable, presented clearly, easily accessible, and introduced at appropriate moments. These measures should be comprehensible and accessible to all young people, regardless of their age, background, or circumstances.
53. Do you agree with the proposed user support measures to be included in the Children's Safety Codes? a) Please confirm which proposed measure your views relate to and provide any arguments and supporting evidence. b) If you responded to our Illegal harms consultation and this is relevant to your response here, please signpost to the relevant parts of your prior response.	Confidential? – N Please see our response to illegal harms.
Search features, functionalities and user support (Section 22)	
54. Do you agree with our proposals? Please provide underlying arguments and evidence to support your views. 55. Do you have additional evidence relating to children's use of search services and the impact of search functionalities on children's behaviour?	Confidential? – Y / N

56. Are there additional steps that you take to protect children from harms as set out in the Act? a) If so, how effective are they? As referenced in the Overview of Codes, Section 13 and Section 17, the use of GenAI to facilitate search is an emerging development and there is currently limited evidence on how the use of GenAI in search services may affect the implementation of the safety measures as set out in this section. We welcome further evidence from stakeholders on the following questions and please provide arguments and evidence to support your views: 57. Do you consider that it is technically feasible to apply the proposed codes measures in respect of GenAI functionalities which are likely to perform or be integrated into search functions? Please provide arguments and evidence to support your views.	

Combined Impact Assessment (Section 23)	
58. Do you agree that our package of proposed measures is proportionate, taking into account the impact on children's safety online as well as the implications on different kinds of services?	Confidential? – Y / N
Statutory tests (Section 24)	
59. Do you agree that our proposals, in particular our proposed recommendations for the draft Children's Safety Codes, are appropriate in the light of the matters to which we must have regard? a) If not, please explain why.	Confidential? – Y / N
Annexes **Impact Assessments (Annex A14)**	
60. In relation to our equality impact assessment, do you agree that some of our proposals would have a positive impact on certain groups? 61. In relation to our Welsh language assessment, do you agree that our proposals are likely to have positive, or more positive impacts on opportunities to use Welsh and treating Welsh no less favourably than English? a) 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.	Confidential? – Y / N

Please complete this form in full and return to protectingchildren@ofcom.org.uk.