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 of the key areas

Safety-by-design

- 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 and large services.
- 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.
- We are also concerned that there is a lack of provision and guidance for multilateral risk such as Sexual Exploitation, which could affect and fall under a series of offences, including: Grooming, Harassment and Financial Crime.
- The omittance of the Blocking Function for smaller services and medium risk large services, is particularly worrying as Childnet's research indicates that children are a lot more likely to block someone, than to proceed with filing an age-inappropriate report.
- Total reports to the Revenge Porn Helpline have seen a tenfold increase in the last four years. While men are predominantly affected by sextortion perpetrated by overseas criminal gangs, women are disproportionately affected by intimate image abuse perpetrated by people known to them. We believe that the guidance does not sufficiently reflect the gendered nature of intimate image abuse.
- As UKSIC, we endorse the suggestion proposed by IWF. While we acknowledge Ofcom's pragmatic approach, we encourage them to exercise greater scrutiny in determining the threshold at which content transitions from being "private" to "publicly" shared. It is imperative for Ofcom to give more thorough consideration to the rights of victims and children, particularly concerning Human Rights and Privacy assessments. This should involve the careful application of principles outlined in Article 8 and 3 of the European Court of Human Rights and Article 19 of the UN Convention on the Rights of the Child. Notably, these principles are not specifically mentioned in Appendix 9 of the guidance.

Accountability and governance

- We feel that the current provisions rely on the "good-will" of the industry as the strategy chosen by OFCOM is reliant on self-assessment of corporations. We would like to see the introduction of an external auditor similar to the process to independently verify approaches.
- We also would like to see increased accountability for 'safety' outcomes placed on industry
- Ofcom has primarily emphasised the size of services rather than assessing their associated risks, leading to the creation of a regulatory framework that exempts numerous services from comprehensive obligations. We contend that the notion "small is safe" is erroneous, and companies with 7 million users should not be considered small;

they are significant entities. We echo the proposal put forward by 5Rights[1] which states that any company with over 2 million UK users should be deemed large. The current size classification excludes significant services such as Fortnite and Roblox with millions of child users, who could be placed in potentially risky and harmful situations

STOP NCII - Datasets

- We agree with the provisions provided to protect children including hash technologies. At the same time, technological developments should be used to protect children as a priority, however, this should not mean that that those developments should not be applied equally to adults where possible.
- StopNCII.org is the world's first device-side hashing technology freely available to any adult in the world to create hashes of their own private sexual content which are then shared with industry partners to prevent the sharing and resharing of that private content on those platforms. Hosted and run by SWGfL, StopNCII.org represents a unique opportunity to protect adults from the non-consensual sharing of intimate images.
- We believe it is essential that Ofcom make it a mandatory requirement for platforms allowing the uploading of user content to take StopNCII.org hashes.

Alternative Dispute Resolution

- We are very concerned that currently users have nowhere to go to challenge decisions made by platforms around the removal of content. We believe that this guidance would benefit greatly from incorporating an alternative dispute resolution model to provide users with the opportunity for redress where content continues to cause harm and is not removed.
- Examples of independent appeals processes exist in Australia and New Zealand, but more countries are now also adopting independent appeals, for example the new Irish Online Safety and Media Regulation Bill 2022 includes provision "for the making of a complaint to the Commission".

UKSIC response – Summary of Key Areas

UKSIC Illegal Harms Response

Volume 2- The causes and impacts of online harm

The importance of Safety-by-design

As the UK Safer Internet Centre, safety online is at the core of our work. On the context of this consultation, the UKSIC would like to acknowledge that the scale of the work of navigating such a complex topic can be really difficult and we would like to acknowledge the efforts of the Ofcom team. Overall, however UKSIC would like to note the fact that the focus of this consultation is industry-centric, which does not reflect the harmful nature and above all the victims of illegal harms online. With that idea in mind, we instead propose a victim-centric approach which will provide a safety-by-design framework, that will facilitate the transition to a safer digital

[1] 5Rights Foundation

environment. The ESafety Commissioner [2] has created a series of principles which accompany the Safety-by-design process which the UKSIC would like to see the inclusion of.

Although we comprehend the approach and classification of risk by size[3], 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. As noted in Volume 2, women and minorities are a lot more likely to face harm and provisions should be put in place to protect them from harm.

UKSIC would also like to reflect on the Illegal Harms Consultation response of SWGfL, which provided data from the Revenge Porn and Report Harmful Content helpline: Whilst volume 2, 6M briefly 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', 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 that affects minorities online.

As evident in volume 1 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. Evidence from We Protect Alliance: Livestreaming - WeProtect Global Alliance.

Adult Illegal Harms

A risk factor not addressed in volume 26L, concerning 'extreme pornography', is the worry that such content may become more hidden or difficult to detect and remove following regulatory measures, making it challenging to address. Currently, the definition of 'extreme pornography' is narrow, yet both the Revenge Porn Helpline and Report Harmful Content services regularly receive reports of content meeting the criteria but falling outside their current scope. This includes content featuring bestiality, rape, and significant violence. For instance, in 2023, there were 30 reports of bestiality-related content. We propose that online platforms sharing content

[2] Principles and background | eSafety Commissioner

[3] Why size and risk matter in our approach to online safety - Ofcom

should be obligated to hash extreme pornographic material to diminish its visibility and thus mitigate harm.

Harmful Content has a significant effect on the user who is exposed to it and around 60% of RPH clients are referred to a mental health service due to significance of impact of their harmful content. UKSIC would therefore propose more effective provisions and regulations which will prevent and mitigate for the harm caused by inappropriate content.

Children Illegal Harms

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.

End-to-End Encryption

We are pleased to note the recognition in Volume 2 (addressing the causes and impacts of online harm) 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[6], 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.

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

Safety-by-design

Reflecting on the publication and statement that was put forward by the OSA network and the supporting organisations, UKSIC feels that the key omittance lies in the fact that although the focus of Ofcom's approach is indeed to build a system that will takedown illegal and harmful content, it does not put into provision the safe-by-design principle. UKSIC comprehends the reasoning behind the size risk approach. However, there should also be consideration for new

[4] OSA Network - Ofcom Illegal Harms - Sign On.docx (onlinesafetyact.net)

[5] Debate: Online Safety Bill - 19th Jul 2023 - Lord Parkinson of Whitley Bay extracts (parallelparliament.co.uk)

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

platforms particularly, with the nature of internet, it provides a ground for start-ups to have a rapid expansion in revenue and user-size.

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]

In most other sectors, evaluating risks associated with a product, feature, or functionality before its introduction and taking measures to mitigate harm are standard practices and fundamental principles of safety by design, as mandated for services already regulated by the UK's Age-Appropriate Design Code[8]. However, Ofcom has opted to only require this type of pre-assessment for the largest services or as a secondary measure in its risk assessment proposals. Proper risk assessments should thoroughly evaluate risks, and new codes that fall below the standards of existing ones not only fail to enhance safety but also risk causing confusion and diluting established best practices.

Governance and Accountability

We believe that Ofcom's suggestions regarding governance and accountability do not go far enough

Ofcom says: "Governance and accountability underpin the way that a service manages risk and ensures that efforts to mitigate them are effective. We consider that these processes are essential components of a well-functioning system of organisational scrutiny, checks and balances, and transparency around risk management activities. Effective governance and accountability processes should be effective in tackling all priority illegal harms." (Volume 3 8.13)

It is promising that many big platforms can point to existing governance structures (and there are likely to be plenty of platforms and smaller services who won't be able to do this). But yet neither we nor Ofcom know the extent to which these existing governance structures are working effectively. For example, Facebook has run into trouble in the past with investors about its oversight structures for risk[9]. This is not a reason to discard what is there of course, but equally Ofcom should not assume that it is sufficient.

We are also quite unsure on how is Ofcom going to assess whether the structures, policies and accountability processes that already exist are sufficient? How will they measure their effectiveness? Is it enough for companies just to say they are doing it? How will they reassess the baseline? There are no examples of how improvements will be measured – either in risk

[7] 5Rights Foundation

[8] Introduction to the Children's code | ICO

[9] Facebook investors demand answers over data scandal (ft.com)

assessment or mitigation (codes). It is also not clear what the difference is between accountability, responsibility and identifiability in relation to governance and senior management roles. Nor is it clear what the written statements of responsibilities will achieve, when Ofcom is primarily citing current practice. E.g. "Several services suggested in their responses to our 2022 Illegal Harms Call for Evidence that they already specify responsibilities for senior members of staff in relation to online safety and risk management".

As noted, there are concerns about the current levels of practice in even the large service providers. Ofcom cites examples of risk assessment best practice, but these are largely focused on reputational risks and external risks to the company, not product safety and design risks created by their own products and services. A product which is accessible for children as young as 13, must protect the users and ensure that the content is age appropriate.

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.

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[10] 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[11], to maintain independence and impartiality.

Adult Illegal Harms

We anticipate significant challenges for Ofcom in ensuring compliance from websites located outside of the UK. Many of the sites we report to, where Non-Consensual Intimate Imagery (NCII) content is shared in a deliberate and harmful manner, are hosted in foreign countries, such as Russia, Malaysia, and South America. This creates considerable obstacles in reporting and removing such content, exacerbating the harmful effects of intimate image abuse, as discussed earlier.

For instance, in Operation Makedom, the Revenge Porn Helpline collaborated with the National Crime Agency to assist around 150 victims affected by a single perpetrator in removing NCII content. Thus far, we have reported over 160,000 individual images and successfully removed over 143,000, achieving a removal rate of 90%. However, many of the remaining 16,000 images are hosted in extensive galleries on dedicated sites, easily accessible to individuals in the UK. Despite the perpetrator being convicted and sentenced to 32 years in prison, the legality of this

[10] Complaint Ex. 1 To Be Sealed MT-IG-AG-NM-000220597 (courtlistener.com)

[11] Our service standards | ICO

content under UK law, as it involves adults, limits our ability to report and remove it. Additionally, it prevents Internet Service Providers (ISPs) from blocking it to decrease visibility and mitigate the outlined harms.

Alternative Dispute Resolution

SWGfL, as a partner in the UK Safer Internet Centre, has been running Report Harmful Content - since 2019. This initiative encourages individuals encountering legally permissible yet harmful content to report it to platforms and offers an independent appeals process. Report Harmful Content (RHC) lacks regulatory authority and instead holds platforms accountable to their own publicly stated terms and conditions.

Data from the 2022 annual report revealed that:

- 11% of reports were elevated to industry platforms, meaning that 11% of the reports submitted to RHC led to an independent appeal process facilitated by us, mediating between a victim and the concerned industry platform. The remaining 89% resulted in further explanations as to why the content did not violate platform community standards.
- Among the reports escalated to industry platforms, 87% were successfully addressed, resulting in the removal of harmful content.
- In approximately one-third of all reports, guidance was provided to direct individuals to the appropriate industry reporting channels.

This data underscores the significance of an independent appeals process within the user reporting system. A considerable number of responses received by victims of harmful content from industry platforms were initially inaccurate, and RHC was able to rectify these situations. Without RHC's intervention, the harm caused may have gone unnoticed or unaddressed.

Amnesty International's[12] research has underscored the exploitation of data-driven, surveillance-oriented business models by tech companies for profit. Through the gathering, retention, and analysis of data, advertisers can target users, including children, steering them towards more extremist content. This exploitation is starkly evident in Myanmar[13], where Facebook's paid advertising tools have been utilized to exacerbate mass violence by disseminating posts that dehumanize and incite violence against the Rohingya community. Furthermore, findings from the Tech Transparency Project[14] indicate that YouTube has been profiting from advertisements featuring white supremacist groups, as well as through the creation of auto-generated "topic" channels, typically reserved for artists with a substantial following, which in some instances could propagate violence-inciting actions.

Alternative Dispute Resolution

Children affected by a service's design features should have avenues for recourse, considering their vulnerability to various online harms. Once a child encounters content or activities that breach a service's legal safety obligations under the Act, prompt reporting and resolution are imperative. However, the Act lacks provisions for individuals to lodge complaints with regulatory authorities or advocacy bodies when they've suffered harm.

[12] https://www.amnesty.org/en/documents/POL40/7349/2023/en/
[13] https://www.amnesty.org/en/documents/asa16/5933/2022/en/

[14] https://www.techtransparencyproject.org/articles/profiting-from-hate-platforms-ad-placement-problem

Existing reporting mechanisms are failing children, particularly:

- Research by the Children's Commissioner[15] for England revealed that 40% of children refrained from reporting harmful content because they believed it would be futile. Additionally, 30% cited a lack of knowledge on how to report, while 25% were unaware that the content could be reported. Only 15% felt that reporting was unnecessary.
- The same research found that platforms often overlook children's reports. Merely 63% of children reported that the content they flagged was removed, while 25% observed no action taken, and 10% were unsure of any outcomes resulting from their reports.

This underscores the need for independent appeals as a component of the Online Safety Act.

UKSIC would also like to share the concern raised by the SWGfL in relation to the recent report from the Public Accounts Committee[16], which highlighted that it could be years before the public saw any demonstrable change in their online lives.

"Ofcom prepared well for its new responsibilities, and moved swiftly to implement the OSA when it became law in October 2023. But the PAC warns of potential public disappointment with the new regulatory regime, which will not be fully implemented until 2026, if people cannot quickly see improvements to their online experience or understand how complaints are acted on. With Ofcom able only to take action where there are systemic concerns about a service provider, the report recommends it develop a mechanism for letting people know what impact their complaint has had".

Dame Meg Hillier MP, Chair of the Committee, said: "*Expectations are understandably high for firm guardrails in the hitherto largely unregulated online world. We know that around two thirds of UK children and adults say they experienced at least one potential online harm in a month in 2022, according to Page 11 Ofcom, which is to be commended for how swiftly it has moved to take on its new responsibilities. It must now continue to be proactively frank with the public over what the Online Safety Act does and does not empower it to do, lest confidence in the new regime be swiftly undermined.*"

"Firm detail on how fees for industry, enforcement, automated monitoring and a range of other issues must now be locked in. No other country has introduced equivalent online safety regulation. Ofcom now needs to capitalise on its early progress. It must also accelerate its coordination with other regulators both at home and overseas, in the recognition that it is at the forefront of a truly global effort to strike the right balance between freedom and safety online."

CSAM

According to the research of IWF[17] and SWGfL[18] a lot of the services that host CSAM or Adult Intimate Abuse Images are hosted abroad, and are operating services in international legal loopholes, where the international policing and Inhope network do not have access to. How can this OSA escalate and assist the process of removal of such content that could even be self-generated by UK citizens.

[15] Digital childhoods: a survey of children and parents | Children's Commissioner for England (childrenscommissioner.gov.uk)

[16] Online Safety Act may take years to have noticeable impact despite public's high expectations - Committees - UK Parliament

[17] Europe remains 'global hub' for hosting of online child sexual abuse material | IWF

[18] Revenge Porn Helpline 2022 Annual Report | SWGfL#

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[19] published by the IWF illustrates that A.I poses a significant risk particularly with the potentially exacerbated volume of csam images 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[20], that should reflect a global online safety regime, where the risk and harm will be minimised.

Smart tools and resources such as Stop-Remove and Stop Ncii, should be encouraged to tackle the exacerbated risks that evolving technologies pose on services and children.

Volume 4: How to mitigate the risk of illegal harms – the illegal content Codes of Practice

Safety-by-design

Risk and size

It's crucial to emphasize that women bear a disproportionate burden of certain online harms, such as harassment, intimate image abuse, and gender-based violence. These aspects warrant more nuanced attention within the Codes of Practice. We recommend incorporating guidance that acknowledges and addresses how online harms disproportionately affect women and girls. Furthermore, it's essential for Ofcom to collaborate with online safety organizations, including SWGfL, which can offer valuable insights into evolving online harms and effective mitigation strategies, particularly those affecting women disproportionately.

The proposed definition of "large services" as those with over 7 million monthly UK users, while straightforward, doesn't fully capture the complexities of online harms. This definition, primarily based on user numbers, overlooks the reality that a platform's size doesn't necessarily correlate directly with the level of harm it may enable. Our experience operating the SWGfL Helplines suggests that some of the most harmful content and behaviours can thrive on smaller platforms. These platforms, due to their size, may lack the scrutiny and oversight applied to larger counterparts, potentially becoming hubs for illegal content and harmful activities.

Moreover, focusing solely on size could create regulatory gaps, disregarding the specific nature and context of illegal content across different platforms. By failing to consider the unique risks posed by the content and the operational and contextual factors of the platform, regulations might not effectively protect users or could inadvertently impose measures on platforms that, despite their large user base, have effective harm mitigation strategies in place.

Age verification and age appropriateness is another significant factor that must be included in the codes of practice. According to the "OFCOM children's media and attitudes report YouTube was the most used site or app among children, visited by 88% of the 3-17-year-olds who go online. This is not surprising, considering that 96% of 3-17-year-olds watch videos online". Particularly worrying is also the fact that "25% of children aged 3-4 used WhatsApp (according to

[19] https://www.iwf.org.uk/about-us/why-we-exist/our-research/how-ai-is-being-abused-to-create-child-sexual-abuse-imagery/

[20] Global collaboration needed as thousands of AI-generated child sexual abuse images emerge depicting the worst kinds of abuse - UK Safer Internet Centre

their parents) compared to 54% of 8-11-year, which are all younger than the age requirement. We therefore strongly believe that Ofcom as the regulatory body should hold companies accountable on the issue of age verification and ensure that content that children access is age appropriate. On the other hand, what is quite positive is that most of the services that children use which appear to be large services such as YouTube, are in scope of this consultation.

Certain children, however, still access video-game platforms. Ofcom in the Children Media Use and Attitude research illustrates that "nine in ten 5-7-year-olds (89%) played video games and a third (34%) played video games online". A lot of these services would be considered as smaller due to the size of the user base but that, user base could equally be of younger age. We propose that because such platforms (who permit messaging and video sharing) pose an increased risk of self-generated CSAM, or grooming, the requirements should be more thorough to mitigate for the risk that this service poses which is unproportional to their size.

Although the Act mandates regulated services to adopt a "proportionate" approach in meeting their obligations, including considering resources, it also requires Ofcom to assess the severity of harm among other factors. We encourage Ofcom to prioritize proportionality concerning the severity and risk of harm to users, especially children, when determining appropriate compliance measures. Unnecessary regulatory burdens can be mitigated through additional support, proportionate mitigation strategies, and regulatory enforcement, as outlined in the Regulators' Code.

Blocking and default settings

In general, UKSIC agrees with the proposals however, we disagree with the "Enhanced User Control" provisions for small services. In particular ref. 9A and 9B which refer to "Users are able to block or mute other individual users and be able to be uncontactable by users they do not yet have an on-service connection with", and "Users can disable comments relating to their own posts, including comments from users that are not blocked".

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 something that Childnet has actively promoted, and is a tool they would use first, ahead of reporting for example, and we have shown this in our own research[21] that it preferrable over reporting, and the logic is that it provides a user control of a situation in a way that reporting doesn't. We would therefore recommend that blocking tools should be a requirement for all services including small services. Childnet's research[22] 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[23]". Blocking is a more effective tool, and the omission of

[21] https://www.childnet.com/blog/young-peoples-views-on-reporting-online-harms/

[22] https://www.childnet.com/blog/young-peoples-views-on-reporting-online-harms/

[23] https://childhub.org/sites/default/files/attachments/Reporting%20Research%20Final.pdf

it in the recommended functions, provides a" fertile" and dangerous ground of grooming, harassment, cyberflashing which could all harm children significantly.

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 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.

Online Content

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[24] 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[25] also conducted research which showcased significance evidence that AI recommend 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[26] provides also insight into the experiences of children online and in particular with recommend 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.

This poses significant dangers in relevance to several priority offences which include Threats Abuse and Harassment and could even escalate to Terrorism offence through the recommend systems that could indoctrinate children into extremist views. It is therefore of great important to

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

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

[26] https://d1xsi6mgo67kia.cloudfront.net/uploads/2024/02/UK-Safer-Internet-Day-2024-Research-Report.pdf

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.

Adult Illegal Harms

Hash Matching and Stop NCII

While prioritizing hash matching for detecting Child Sexual Abuse Material (CSAM) is crucial and we would like to applaud the efforts of the OFCOM team, we think that the technology and responsibility should extend to address other forms of illegal and harmful content, such as terrorism, Non-Consensual Intimate Images (NCII), and extreme pornography. This broader application recognises the diverse nature of online harms and ensures a more comprehensive approach to protecting users. Similarly, the reporting and complaints section, currently covering 'CSEA, Terrorism, and Other duties,' should explicitly include responsibilities related to NCII, extreme pornography, and other significant harms. This specificity will provide platforms with clear guidelines on the range of content requiring vigilant monitoring and response, thereby closing potential loopholes that could leave users vulnerable to harm. These improvements would not only enhance the clarity and efficacy of the Codes but also demonstrate a deeper understanding of the online risk users face, fostering a safer internet environment for everyone.

While it's imperative to prioritize technological advancements to protect children, this shouldn't imply that such developments shouldn't be equally applied to adults wherever feasible.

In 2021, SWGfL collaborated with Meta to develop the StopNCII.org platform, allowing adults to generate hashes of their intimate images to prevent them from being shared without consent on participating platforms. Currently, StopNCII.org safeguards over 500,000 individual images from being shared across nine participating platforms, including Facebook, Instagram, Threads, Reddit, Bumble, TikTok, OnlyFans, Aylo (formerly MindGeek inc PornHub), and Snap[27]. We've actively prevented over 11,000 NCII images from being shared.

In our experience, the most harmful content can appear on the smallest platforms. By excluding small platforms from the most onerous measures, it removes oversight of the riskiest environments and leaves opportunity for harm to occur unchecked.

Hash matching technology which is on the context of this consultation solely used for the removal of CSAM, should also be expanded to include other forms of illegal content most notably A.11 Adult image-based sexual offences, and non-priority offences. Tools such as StopNcii which is operated by SWGfL (A partner at the UK Safer Internet Centre) should be encouraged and used by smaller services with high or medium risk functions such as messaging, URL sharing and video-sharing platforms.

Children Illegal Harms

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

[27] How StopNCII.org Works | StopNCII.org

[28] 5Rights Foundation

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.

The Age-Appropriate Design Code (AADC)[29], which outlines the treatment of children's data by services within its scope, combines both outcomes-based and prescriptive standards. Since its introduction, major technology companies like Google, TikTok, and Meta have implemented numerous design adjustments, such as setting children's accounts to private by default, disabling notifications and direct messaging, and ensuring transparency, to meet the code's requirements. The AADC has spurred innovation among tech firms to enhance online safety and improve children's online experiences.

However, we express concern regarding the absence of outcomes-based standards in the code, which contradicts its stated objectives. During the Act's passage, Lord Minister Parkinson of Whitley Bay emphasized that the codes should be outcomes-based and not overly prescriptive, as this could hinder smaller services' ability to comply. He stated, "We must also acknowledge the diversity and innovative nature of this sector. Requiring compliance with specific steps rather than focusing on outcomes might result in companies not employing the most effective or efficient methods to safeguard children."

CSAM

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 a 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[30] 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

[29] age-appropriate-design-a-code-of-practice-for-online-services-2-1.pdf (ico.org.uk)

[30] https://www.lucyfaithfull.org.uk/files/reThink_Chatbot_Evaluation_Report.pdf

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 behavior 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.

Sexual exploitation

We would also like to make a note of the fact that Sextortion appears to be missing from the Grooming, and CSAM priority offences list, which are both set out in Annex 10. Our worry reflects also a greater risk regarding the classification of risk particularly with offences such as extortion which could fall under multiple priority offences such as Harassment, Grooming and Sexual Exploitation of Adults. The UKSIC has hosted an Insight Research Series[31] 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"[32].

Complicated offences such as sexual exploitation which could impact a multitude of other offences[33] and encompass an array of maleficent actors, what would be the provisions for the inclusion of a multifaceted risk in the risk registry?

What could also be linked to sexual exploitation of children, is the proposed guidance for verification schemes. According to user Control Ref.9C[34] we disagree with the lack of requirements set out for smaller services for internal and transparent policies regarding their verification and paid-for-verification scheme. Implementing a verification scheme can be useful,

[31] UK Safer Internet Centre UK Safer Internet Centre - UK Safer Internet Centre

[32] Hotline reports 'shocking' rise in the sextortion of boys (iwf.org.uk)

[33] https://www.gov.uk/government/publications/tackling-child-sexual-abuse-strategy

[34] https://www.ofcom.org.uk/__data/assets/pdf_file/0023/271175/Consultation-at-a-glance-our-proposals-and-who-they-apply-to.pdf

especially with issues like misinformation which are mentioned in volume 2. However, the lack of a transparent process which is not a requirement for most services could lead to exacerbated danger for children and could amplify risk for issues such as grooming and sextortion (Understanding Verification on Instagram). According to the BBC Bitesize research[35], 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[36] indicates that nearly a quarter (23%) of 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.

Recommendations

Firstly, we propose a couple of adjustments to the current codes in line with the proposals put forward by IWF:

- We advocate for the immediate inclusion of **classifier technology** for detecting new instances of child sexual abuse as a mitigation measure. Its absence represents a significant oversight, especially considering that many of the large services covered by this regulation are already employing such technology.
- Secondly, we recommend the addition of **Age Verification measures** to enhance the effectiveness of grooming mitigations without delay. Relying solely on self-declaration of age for children's accounts is insufficient, particularly when considering that Ofcom is currently soliciting input on Age Assurance as part of its obligations under Part 5 of the Act.
- Lastly, we suggest that Ofcom should endorse the utilization of keyword databases for both **User-to-User and Search Services**, leveraging available services provided by organizations like ours. Keyword detection has been proposed by Ofcom as a means to mitigate fraud, and as a technology, it should be straightforward to demonstrate its compliance with accuracy, effectiveness, and bias-free requirements.

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

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

Volume 5: How to judge whether content is illegal or not? (Illegal Content Judgements Guidance) & Volume 6: Information gathering and enforcement powers and approach to supervision

Safety-by-design

Ofcom deserves commendation for swiftly assuming its new responsibilities. However, it must maintain transparent communication with the public about the scope of authority granted by the Online Safety Act to prevent erosion of confidence in the new regulatory framework.

Concrete plans regarding industry fees, enforcement procedures, automated monitoring, and other pertinent matters need to be promptly established. Considering that no other country has implemented comparable online safety regulations, Ofcom must seize the opportunity to build on its initial progress. Moreover, it should expedite collaboration with domestic and international regulators, recognizing its pivotal role in leading a global initiative to find the appropriate equilibrium between online freedom and safety.

Support the outlined default settings and can see the protections these will offer young people. And support the provision of supportive information in a timely and accessible manner to help users make informed choices when they seek to change their settings, for e.g. to disable default, or receive a direct message from another user for the first time. These measures support wider digital literacy as well as provide key potential protections against risks such as grooming and financial online sextortion.

Age Verification

We express concern alongside IWF regarding the grooming mitigations outlined on pages 229 and 230 of volume 4, as they currently rely solely on self-declared age, which can be easily manipulated by children simply by providing false information during registration.

While we acknowledge that Ofcom intends to address the issue of Age Verification through the forthcoming "protection of children" code expected in the coming months, we believe it is illogical to propose measures that claim to significantly impact grooming while being susceptible to such easy circumvention. This is especially concerning given the well-established Age Verification industry and Ofcom's ongoing consultation on Age Assurance measures as part of the provisions under Part 5 of the implementation of the Online Safety Act.

We recommend that Ofcom incorporates Age Verification measures alongside the Grooming measures to enhance their effectiveness.

Adult Illegal Harms

We also agree with the inclusion of election interfering as a relevant priority offence Foreign interference and false communications, particularly with the political sensitive election periods which targeted harmed b target voters with Deepfakes and the spread of misinformation. A call for international cooperation would also be really important.

Stop Ncii

We express our support for the measures addressing 'Adult image-based sexual offences' highlighted in volume 5, and we welcome the reinforcement of intimate image abuse laws in England and Wales. However, we observe a notable oversight regarding the responsibility of service providers to remove non-consensually shared intimate content that has been reshared after the initial offence. The current guidance's position, which suggests no mandatory action for

service providers upon notification of non-consensual image sharing, is worrisome. Relying solely on platforms' voluntary compliance is inadequate, given the historical evidence that delayed or inconsistent responses significantly impact adults affected by intimate image abuse. It is essential that platforms are strongly compelled to promptly remove any known non-consensual images to effectively prevent further harm. The continuous sharing and resharing of content exacerbate the harm inflicted on the victim. Platforms whose business models rely on such content attract motivated users who persist in resharing and downloading, perpetuating sustained harm to victims. Multiple resharing instances of Non-Consensual Intimate Images (NCII) content amplify the trauma, and platforms should be mandated to remove subsequently shared content as swiftly as the initial one.

Furthermore, the guidance should consider the recommended integration of the StopNCII.org platform, operated by SWGfL, as discussed in our response to volume 4, to bolster content detection processes. While the importance of hashing and detecting Child Sexual Abuse Material (CSAM) is emphasized in the guidance, similar considerations for other priority offence content, such as non-consensually shared intimate images, should also be given substantial attention.

Children Illegal Harms

It is encouraging that, in accordance with the law, the choice to pursue enforcement measures will focus on instances where the service has violated its obligations regarding child safety.

Concerning paragraph 29.39(b), when assessing whether children can access certain parts or the entirety of the service, Ofcom must verify that the age verification methods implemented by the service adhere to the standards outlined in the age assurance guidance. Merely having age verification mechanisms in place does not guarantee that children cannot access the service; the effectiveness of these measures depends on their quality. The enforcement process should act as a mechanism that will protect children from accessing harmful and age-inappropriate content.

Various international examples of exemplary practices exist regarding the integration of safety by design. For instance, the Australian e-safety commissioner[37] has formulated principles, an accessible assessment tool for services, resources tailored for investors and financial entities, and guidance for the tertiary sector on effectively engaging all relevant components of a safety-by-design process. We encourage Ofcom to contemplate a similar strategy for regulation in the UK.

Although there is some indication of adherence to safety-by-design principles, such as the proposed Grooming mitigations in Ofcom's code of practice, these measures are presently limited to the largest platforms or those deemed to be at medium to high risk of Child Sexual Abuse Material (CSAM).

We perceive it as a missed opportunity not to capitalize on the successes of implementing the Age-Appropriate Design Code[38] to ensure that platforms incorporate safety measures from the outset, rather than constantly having to retrofit solutions to combat the spread of illegal content on their platforms.

[37] https://www.esafety.gov.au/industry/safety-by-design

[38] Introduction to the Children's code | ICO

General Comments

Alternative Dispute Resolution

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 can be found in volume 3 above. 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[39] 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.

Risk and Size

We agree with a part of the principle as indeed user size, could to a certain extent lead to a higher risk, on the account of the user size of the platform. Nevertheless, we believe that the most significant indicator besides size should be instead the risk, which is drawn by the functions of the platform. We are therefore proposing a victim-centric approach to illegal harms, which will entail that services of any size must protect their users from illegal harms. A safety-by-design principle is not applied on this occasion, particularly as the key reasoning for the Size risk classification, lies in the efforts of Ofcom to limit the financial costs to companies[40]. In our view, the focus should be placed on the victims and the users of the platform.

The user size for a company to be considered is quite high so a lot of significant services with a large user size would not be in scope of the large-service recommendations, which are more comprehensive. Reflecting on the analysis from the Online Safety Act Network, companies such

[39] Report Harmful Content Release Final Quarterly Report for 2021 | SWGfL

[40] https://www.ofcom.org.uk/news-centre/2024/why-size-and-risk-matter-in-our-approach-to-online-safety

as Roblox and Fortnite would not be classified as a large service and could place millions of children at risk.

Safety-by-design

In general, however we disagree with the industry-centric approach of this consultation, instead we would prefer to witness and contribute to the adoption of 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[41], there is clear evidence that children are affected negatively when using such platforms: (36%) of children notice a negative change in their mental change 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. Recommend 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.

We also urge 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.

According to Schedule 4, Paragraph 13 (6), Ofcom is required to consider the accuracy, effectiveness, and impartiality of the technology employed. While Ofcom has effectively outlined how Hash Matching and URL blocking meet these criteria, it would also be beneficial for Ofcom to adopt a similar approach when recommending technologies for enforcement actions under Section 122 of the Bill, particularly in relation to the Use of Technology Notices, which are yet to be developed.

[41] https://d1xsi6mgo67kia.cloudfront.net/uploads/2024/02/UK-Safer-Internet-Day-2024-Research-Report.pdf

Your response

Question (Volume 2)	Your response
Question 6.1: Do you have any comments on Ofcom's assessment of the causes and impacts of online harms? Do you think we have missed anything important in our analysis? Please provide evidence to support your answer.	*[Is this answer confidential? Yes / No (delete as appropriate)]* Safety by design UKSIC would also like to reflect on the Illegal Harms Consultation response of SWGfL, which provided data from the Revenge Porn and Report Harmful Content helpline: Whilst volume 2, 6M briefly 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,

Question (Volume 2)	Your response
	TikTok, and YouTube were reported for removal by Report Harmful Content" <u>Generative Artificial Intelligence:</u> In Volume 2, Page 6, point 5.6, Ofcom states its intention to monitor emerging risks and trends in regulated services and update its Risk Register accordingly. This may involve expanding the scope of risk assessment to include technologies like immersive online virtual worlds, augmented realities, and generative artificial intelligence ('generative AI'). However, the footnote clarifies that the Risk Register has only partially addressed the risk of Generative AI technologies, considering "some of these risks." In another section discussing the risks of Child Sexual Abuse Material (CSAM) in Volume 2, the consultation notes the challenge posed by deepfakes, stating that their identification is difficult. We argue that Generative AI is not merely a future risk but requires immediate oversight to prevent exacerbation of the problem. In a recent one-month period between September and October 2023, the Internet Watch Foundation (IWF) discovered 20,254 AI-generated images on a dark web forum. Upon human review of 11,108 images, 2,978 were found to be illegal under existing legislation. Ahead of the UK Government's International AI Safety Summit, the IWF issued a report containing recommendations on regulating the technology to prevent harm. These recommendations include: Ensuring that data sets used for generative AI undergo scrutiny by expert child safety organizations to ensure they are free of child sexual abuse material. Enforcing protections in closed-source models and subjecting open-source models to regulatory review to implement appropriate risk mitigation measures. Requiring companies to establish clear terms and conditions prohibiting users from generating CSAM with their tools. Instructing search providers to de-index finely tuned models associated with the creation of AI-generated CSAM. Furthermore, discussions have been ongoing regarding additional mitigations, such as holding app stores accountable

Question (Volume 2)	Your response
	for removing applications known to be used for generating AI CSAM in violation of their terms and conditions.
Question 6.2: 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.	*[Is this answer confidential? Yes / No (delete as appropriate)]* <u>Safety by Design:</u> Additionally, the 'Digital Misogynoir Report: Ending the dehumanising of Black women on social media', 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 that affects minorities online. As evident in volume 1 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. Evidence from We Protect Alliance: Livestreaming - WeProtect Global Alliance. **Children Illegal Harms** 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[42] 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" [43]. Safety and innovation can co-exist, and the regulation and

[42] OSA Network - Ofcom Illegal Harms - Sign On.docx (onlinesafetyact.net)

[43] Debate: Online Safety Bill - 19th Jul 2023 - Lord Parkinson of Whitley Bay extracts (parallelparliament.co.uk)

Question (Volume 2)	Your response
	processes must keep their users safe and most importantly vulnerable groups such as children. End-to-End Encryption (E2EE): We are encouraged by the recognition of End-to-End Encryption as a feature posing specific risks, especially concerning the detection of Child Sexual Exploitation and Abuse (CSE/A). When considering this aspect alongside other characteristics of a service and its potential involvement with child sexual abuse material, we anticipate a notable impact on harm to children. As highlighted in this consultation, the instance of Meta's suspension of scanning for child sexual abuse content on EU-based accounts resulted in a significant decrease in reports to the National Center for Missing and Exploited Children (NCMEC) by 58%. Should they implement encryption for their messenger and Instagram direct messaging features, similar declines in reporting may occur. This effect could potentially be more pronounced since End-to-End Encryption will be implemented globally, not solely for EU accounts. Measurement of service size: Page 11 of Volume 2 outlines two methods proposed by Ofcom to assess service size: measuring the user base and the number of employees (capacity). These metrics resemble the criteria used to determine membership fees for the Internet Watch Foundation (IWF), which are based on size, as determined by Ofcom's outlined criteria, and sector. However, in subsequent volumes of the consultation (Volume 3 and Chapter 11), Ofcom indicates that a "large" service will be defined as one with a user base exceeding 7 million monthly UK users. Our concern is that this approach might overlook many websites responsible for hosting significant amounts of child sexual abuse material. Of particular concern is the differing approaches to risk assessment.

Question (Volume 3)	Your response
Question 8.1: Do you agree with our proposals in relation to governance and accountability measures in the illegal content Codes of Practice? Please provide underlying arguments and evidence of efficacy or risks to support your view.	*[Is this answer confidential? No* Governance and Accountability We believe that Ofcom's suggestions regarding governance and accountability do not go far enough Ofcom says: "Governance and accountability underpin the way that a service manages risk and ensures that efforts to mitigate them are effective. We consider that these processes are essential components of a well-functioning system of organisational scrutiny, checks and balances, and transparency around risk management activities. Effective governance and accountability processes should be effective in tackling all priority illegal harms." (Volume 3 8.13) It is promising that many big platforms can point to existing governance structures (and there are likely to be plenty of platforms and smaller services who won't be able to do this). But yet neither we nor Ofcom know the extent to which these existing governance structures are working effectively. For example, Facebook has run into trouble in the past with investors about its oversight structures for risk[44]. This is not a reason to discard what is there of course, but equally Ofcom should not assume that it is sufficient. We are also quite unsure on how is Ofcom going to assess whether the structures, policies and accountability processes that already exist are sufficient? How will they measure their effectiveness? Is it enough for companies just to say they are doing it? How will they reassess the baseline? There are no examples of how improvements will be measured – either in risk assessment or mitigation (codes). It is also not clear what the difference is between accountability, responsibility and identifiability in relation to governance and senior management roles. Nor is it clear what the written statements of responsibilities will achieve, when Ofcom is primarily citing current practice. E.g. "Several services suggested in their responses to our 2022 Illegal Harms Call for Evidence that they

[44] Facebook investors demand answers over data scandal (ft.com)

Question (Volume 3)	Your response
	already specify responsibilities for senior members of staff in relation to online safety and risk management". As noted, there are concerns about the current levels of practice in even the large service providers. Ofcom cites examples of risk assessment best practice, but these are largely focused on reputational risks and external risks to the company, not product safety and design risks created by their own products and services. A product which is accessible for children as young as 13, must protect the users and ensure that the content is age appropriate. 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. 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[45] 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[46], to maintain independence and impartiality.

[45] Complaint Ex. 1 To Be Sealed MT-IG-AG-NM-000220597 (courtlistener.com)

[46] Our service standards | ICO

Question (Volume 3)	Your response
Question 8.2: Do you agree with the types of services that we propose the governance and accountability measures should apply to?	*Is this answer confidential? No* Online safety measures should be tailored to the risks inherent in the service, taking into account factors such as subject matter and functionality, regardless of the service's size, intentions, or potential expenses. However, this consultation overly emphasizes the financial burden on services rather than the severity of the harm they pose. Such a focus may incentivize companies to shift the costs of addressing harm to the public sector, including health, justice, and education services. Moreover, it could widen the competitive gap, favoring larger services that can better absorb these costs. This is especially troubling when considering content or activities that are deemed illegal. <u>Risk and size</u> It's crucial to emphasize that women bear a disproportionate burden of certain online harms, such as harassment, intimate image abuse, and gender-based violence. These aspects warrant more nuanced attention within the Codes of Practice. We recommend incorporating guidance that acknowledges and addresses how online harms disproportionately affect women and girls. Furthermore, it's essential for Ofcom to collaborate with online safety organizations, including SWGfL, which can offer valuable insights into evolving online harms and effective mitigation strategies, particularly those affecting women disproportionately. The proposed definition of "large services" as those with over 7 million monthly UK users, while straightforward, doesn't fully capture the complexities of online harms. This definition, primarily based on user numbers, overlooks the reality that a platform's size doesn't necessarily correlate directly with the level of harm it may enable. Our experience operating the SWGfL Helplines suggests that some of the most harmful content and behaviours can thrive on smaller platforms. These platforms, due to their size, may lack the scrutiny and oversight applied to larger counterparts, potentially becoming hubs for illegal content and harmful activities. Moreover, focusing solely on size could create regulatory gaps, disregarding the specific nature and

Question (Volume 3)	Your response
	context of illegal content across different platforms. By failing to consider the unique risks posed by the content and the operational and contextual factors of the platform, regulations might not effectively protect users or could inadvertently impose measures on platforms that, despite their large user base, have effective harm mitigation strategies in place.
Question 8.3: 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?	*Is this answer confidential? No* **Adult Illegal Harms** We anticipate significant challenges for Ofcom in ensuring compliance from websites located outside of the UK. Many of the sites we report to, where Non-Consensual Intimate Imagery (NCII) content is shared in a deliberate and harmful manner, are hosted in foreign countries, such as Russia, Malaysia, and South America. This creates considerable obstacles in reporting and removing such content, exacerbating the harmful effects of intimate image abuse, as discussed earlier. For instance, in Operation Makedom, the Revenge Porn Helpline collaborated with the National Crime Agency to assist around 150 victims affected by a single perpetrator in removing NCII content. Thus far, we have reported over 160,000 individual images and successfully removed over 143,000, achieving a removal rate of 90%. However, many of the remaining 16,000 images are hosted in extensive galleries on dedicated sites, easily accessible to individuals in the UK. Despite the perpetrator being convicted and sentenced to 32 years in prison, the legality of this content under UK law, as it involves adults, limits our ability to report and remove it. Additionally, it prevents Internet Service Providers (ISPs) from blocking it to decrease visibility and mitigate the outlined harms. Alternative Dispute Resolution SWGfL, as a partner in the UK Safer Internet Centre, has been running Report Harmful Content - since 2019. This initiative encourages individuals encountering legally permissible yet harmful content to report it to platforms and offers an independent appeals process. Report Harmful Content (RHC) lacks regulatory authority and

Question (Volume 3)	Your response
	instead holds platforms accountable to their own publicly stated terms and conditions. Data from the 2022 annual report revealed that: • 11% of reports were elevated to industry platforms, meaning that 11% of the reports submitted to RHC led to an independent appeal process facilitated by us, mediating between a victim and the concerned industry platform. The remaining 89% resulted in further explanations as to why the content did not violate platform community standards. • Among the reports escalated to industry platforms, 87% were successfully addressed, resulting in the removal of harmful content. • In approximately one-third of all reports, guidance was provided to direct individuals to the appropriate industry reporting channels. This data underscores the significance of an independent appeals process within the user reporting system. A considerable number of responses received by victims of harmful content from industry platforms were initially inaccurate, and RHC was able to rectify these situations. Without RHC's intervention, the harm caused may have gone unnoticed or unaddressed.
Question: 8.4: 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?	*Is this answer confidential? No* We are not entirely convinced that a self-appraisal and assessment model will be particularly effective in an industry that had significant issues in the past with governance and accountability. Most recently, Arturo Bejar's testimony[47] shed a light into the operations and safety measures of the tech industry, which have caused harm to millions of users.

[47] https://www.judiciary.senate.gov/imo/media/doc/2023-11-07_-_testimony_-_bejar.pdf

Question (Volume 3)	Your response
Question 9.1: Do you agree with our proposals? Please provide the underlying arguments and evidence that support your views.	*Is this answer confidential? No* <u>Safety-By-Design</u> Although we comprehend the approach and classification of risk by size[3], 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. As noted in Volume 2, women and minorities are a lot more likely to face harm and provisions should be put in place to protect them from harm. **Children Illegal Harms** 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. **Recommendations:** We have set out above where we disagree with proposals. To summarise we would like to see: • The definition of Very Large Platforms revisited to ensure more services are caught in scope of the regulation.

Question (Volume 3)	Your response
	• We believe governance and accountability measures should apply to services at medium to high risk of one harm. • Training requirements should be extended to staff in services where they are deemed to be medium to high risk of CSAM. Reflecting on the publication and statement that was put forward by the OSA network and the supporting organizations, UKSIC feels that the key omittance lies in the fact that although the focus of Ofcom's approach is indeed to build a system that will takedown illegal and harmful content, it does not put into provision the safe-by-design principle. UKSIC comprehends the reasoning behind the size risk approach however, there should also be consideration for new platforms particularly, with the nature of internet, it provides a ground for start-ups to have a rapid expansion in revenue and user-size.
Question 9.2: 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?	No comment
Question 9.3: Are the Risk Profiles sufficiently clear and do you think the information provided on risk factors will help you understand the risks on your service?[48]	*Is this answer confidential? No* <u>Safety-by-design</u> In general, however we disagree with the industry-centric approach of this consultation, instead we would prefer to witness and contribute to the adoption of 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[41], there is clear evidence that children are affected negatively when using such platforms: (36%) of children notice a

[48] If you have comments or input related the links between different kinds of illegal harm and risk factors, please refer to Volume 2: Chapter 5 Summary of the causes and impacts of online harm).

Question (Volume 3)	Your response
	negative change in their mental change 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. Recommend 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. We also urge 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. According to Schedule 4, Paragraph 13 (6), Ofcom is required to consider the accuracy, effectiveness, and impartiality of the technology employed. While Ofcom has effectively outlined how Hash Matching and URL blocking meet these criteria, it would also be beneficial for Ofcom to adopt a similar approach when recommending technologies for enforcement actions under Section 122 of the Bill, particularly in relation to the Use of Technology Notices, which are yet to be developed.
Question 10.1: Do you have any comments on our draft record keeping and review guidance?	No Comment

Question (Volume 3)	Your response
Question 10.2: 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?	No Comment

Question (Volume 4)	Your response
Question 11.1: Do you have any comments on our overarching approach to developing our illegal content Codes of Practice?	The proposed definition of "large services" as those with over 7 million monthly UK users, while straightforward, doesn't fully capture the complexities of online harms. This definition, primarily based on user numbers, overlooks the reality that a platform's size doesn't necessarily correlate directly with the level of harm it may enable. Our experience operating the SWGfL Helplines suggests that some of the most harmful content and behaviours can thrive on smaller platforms. These platforms, due to their size, may lack the scrutiny and oversight applied to larger counterparts, potentially becoming hubs for illegal content and harmful activities. Moreover, focusing solely on size could create regulatory gaps, disregarding the specific nature and context of illegal content across different platforms. By failing to consider the unique risks posed by the content and the operational and contextual factors of the platform, regulations might not effectively protect users or could inadvertently impose measures on platforms that, despite their large user base, have effective harm mitigation strategies in place. Age verification and age appropriateness is another significant factor that must be included in the codes of practice. According to the "OFCOM children's media and attitudes report YouTube was the most used site or app among children, visited by 88% of the 3-17-year-olds who go online. This is not surprising, considering that 96% of 3-17-year-olds watch videos online". Particularly worrying is also the fact that "25% of children aged 3-4 used WhatsApp (according to their parents) compared to 54% of 8-11-year, which are all younger than the age

Question (Volume 4)	Your response
	requirement. We therefore strongly believe that Ofcom as the regulatory body should hold companies accountable on the issue of age verification and ensure that content that children access is age appropriate. On the other hand, what is quite positive is that most of the services that children use which appear to be large services such as YouTube, are in scope of this consultation.
Question 11.2: 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?	*Is this answer confidential? No* Risk and Size We agree with a part of the principle as indeed user size, could to a certain extent lead to a higher risk, on the account of the user size of the platform. Nevertheless, we believe that the most significant indicator besides size should be instead the risk, which is drawn by the functions of the platform. We are therefore proposing a victim-centric approach to illegal harms, which will entail that services of any size must protect their users from illegal harms. A safety-by-design principle is not applied on this occasion, particularly as the key reasoning for the Size risk classification, lies in the efforts of Ofcom to limit the financial costs to companies[40]. In our view, the focus should be placed on the victims and the users of the platform. The user size for a company to be considered is quite high so a lot of significant services with a large user size would not be in scope of the large-service recommendations, which are more comprehensive. Reflecting on the analysis from the Online Safety Act Network, companies such as Roblox and Fortnite would not be classified as a large service and could place millions of children at risk. **Children Illegal Harms** 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

Question (Volume 4)	Your response
	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. The Age-Appropriate Design Code (AADC)[29], which outlines the treatment of children's data by services within its scope, combines both outcomes-based and prescriptive standards. Since its introduction, major technology companies like Google, TikTok, and Meta have implemented numerous design adjustments, such as setting children's accounts to private by default, disabling notifications and direct messaging, and ensuring transparency, to meet the code's requirements. The AADC has spurred innovation among tech firms to enhance online safety and improve children's online experiences. However, we express concern regarding the absence of outcomes-based standards in the code, which contradicts its stated objectives. During the Act's passage, Lord Minister Parkinson of Whitley Bay emphasized that the codes should be outcomes-based and not overly prescriptive, as this could hinder smaller services' ability to comply. He stated, "We must also acknowledge the diversity and innovative nature of this sector. Requiring compliance with specific steps rather than focusing on outcomes might result in companies not employing the most effective or efficient methods to safeguard children."
Question 11.3: Do you agree with our definition of large services?	*Is this answer confidential? No* **Children Illegal Harms** 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

Question (Volume 4)	Your response
	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.
Question 11.4: Do you agree with our definition of multi-risk services?	*Is this answer confidential? No* Sexual exploitation We would also like to make a note of the fact that Sextortion appears to be missing from the Grooming, and CSAM priority offences list, which are both set out in Annex 10. Our worry reflects also a greater risk regarding the classification of risk particularly with offences such as extortion which could fall under multiple priority offences such as Harassment, Grooming and Sexual Exploitation of Adults. The UKSIC has hosted an Insight Research Series[31] 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,

Question (Volume 4)	Your response
	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"[32]. Complicated offences such as sexual exploitation which could impact a multitude of other offences[33] and encompass an array of maleficent actors, what would be the provisions for the inclusion of a multifaceted risk in the risk registry? What could also be linked to sexual exploitation of children, is the proposed guidance for verification schemes. According to user Control Ref.9C[34] we disagree with the lack of requirements set out for smaller services for internal and transparent policies regarding their verification and paid-for-verification scheme. Implementing a verification scheme can be useful, especially with issues like misinformation which are mentioned in volume 2. However, the lack of a transparent process which is not a requirement for most services could lead to exacerbated danger for children and could amplify risk for issues such as grooming and sextortion (Understanding Verification on Instagram). According to the BBC Bitesize research[35], 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.
Question 11.6: Do you have any comments on the draft Codes of Practice themselves?[49]	*Is this answer confidential? No* Blocking and default settings In general, UKSIC agrees with the proposals however, we disagree with the "Enhanced User Control" provisions for small services. In particular ref. 9A and 9B which refer to "Users are able to block or mute other individual users and be able to be uncontactable by users they do not yet have an on-service connection with", and "Users can

[49] See Annexes 7 and 8.

Question (Volume 4)	Your response
	disable comments relating to their own posts, including comments from users that are not blocked". 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 something that Childnet has actively promoted, and is a tool they would use first, ahead of reporting for example, and we have shown this in our own research[21] that it preferrable over reporting, and the logic is that it provides a user control of a situation in a way that reporting doesn't. We would therefore recommend that blocking tools should be a requirement for all services including small services. Childnet's research[22] 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[23]". 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. 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

Question (Volume 4)	Your response
	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 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. Reflecting also on the IWF response, we recommend the following measures are added: • Keyword detection for CSAM • Use of classifiers (AI and Machine Learning) to detect CSAM content that has not previously been detected • Grooming measures are supported by Age Verification and not reliant on self-declaration of age. • Codes of Practice are amended to require companies to mitigate risks identified in their risk assessment.
Question 11.7: Do you have any comments on the costs assumptions set out in Annex 14, which we used for calculating the costs of various measures?	Nothing further to add
Question 12.1: Do you agree with our proposals? Please provide the underlying arguments and evidence that support your views.	*Is this answer confidential? No* <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 a good practice which combines an automated and manual content moderation with an effective process which includes hash/matching, URL

Question (Volume 4)	Your response
	matching and a cross industry keyword list, could all contribute to a more effective content regulation. In February 2024, a study[30] 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 behavior 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

Question (Volume 4)	Your response
	lack of evidence, especially when it is already considered best practice within the industry.
Question 13.1: Do you agree with our proposals? Please provide the underlying arguments and evidence that support your views.	*Is this answer confidential? No* <u>End-to-End encryption</u> We are pleased to note the recognition in Volume 2 (addressing the causes and impacts of online harm) 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[6], 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.
Question 14.1: Do you agree with our proposals? Do you have any views on our three proposals, i.e. CSAM hash matching, CSAM URL detection and fraud keyword detection? Please provide the underlying arguments and evidence that support your views.	*Is this answer confidential? No* <u>CSAM</u> According to the research of IWF[17] and SWGfL[18] a lot of the services that host CSAM or Adult Intimate Abuse Images are hosted abroad, and are operating services in international legal loopholes, where the international policing and Inhope network do not have access to. How can this OSA escalate and assist the process of removal of such content that could even be self-generated by UK citizens. 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[19] published by the IWF illustrates that A.I poses a significant risk particularly with the potentially exacerbated volume of CSAM images 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

Question (Volume 4)	Your response
	as" small" due to their user size. UKSIC would therefore agree with the call of global cooperation that IWF proposed in 2023[20], that should reflect a global online safety regime, where the risk and harm will be minimised. Smart tools and resources such as Stop-Remove and Stop Ncii, should be encouraged to tackle the exacerbated risks that evolving technologies pose on services and children.
Question 14.2: Do you have any comments on the draft guidance set out in Annex 9 regarding whether content is communicated 'publicly' or 'privately'?	No Comment
Question 14.3: Do you have any relevant evidence on: • The accuracy of perceptual hash matching and the costs of applying CSAM hash matching to smaller services; • The 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; • The costs of applying our CSAM URL detection measure to smaller services, and the effectiveness of fuzzy	*Is this answer confidential? No* <u>Hash Matching and Stop NCII</u> While prioritizing hash matching for detecting Child Sexual Abuse Material (CSAM) is crucial and we would like to applaud the efforts of the OFCOM team, we think that the technology and responsibility should extend to address other forms of illegal and harmful content, such as terrorism, Non-Consensual Intimate Images (NCII), and extreme pornography. This broader application recognises the diverse nature of online harms and ensures a more comprehensive approach to protecting users. Similarly, the reporting and complaints section, currently covering 'CSEA, Terrorism, and Other duties,' should explicitly include responsibilities related to NCII, extreme pornography, and other significant harms. This specificity will provide platforms with clear guidelines on the range of content requiring vigilant monitoring and response, thereby closing potential loopholes that could leave users vulnerable to harm. These improvements would not only enhance the clarity and efficacy of the Codes but also demonstrate a deeper understanding of the online risk users face, fostering a safer internet environment for everyone. While it's imperative to prioritize technological advancements to protect children, this shouldn't imply that such developments shouldn't be equally applied to adults wherever feasible.

Question (Volume 4)	Your response
matching[50] for CSAM URL detection; • The costs of applying our articles for use in frauds (standard keyword detection) measure, including for smaller services; and • 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.	In 2021, SWGfL collaborated with Meta to develop the StopNCII.org platform, allowing adults to generate hashes of their intimate images to prevent them from being shared without consent on participating platforms. Currently, StopNCII.org safeguards over 500,000 individual images from being shared across nine participating platforms, including Facebook, Instagram, Threads, Reddit, Bumble, TikTok, OnlyFans, Aylo (formerly MindGeek inc PornHub), and Snap27. We've actively prevented over 11,000 NCII images from being shared. In our experience, the most harmful content can appear on the smallest platforms. By excluding small platforms from the most onerous measures, it removes oversight of the riskiest environments and leaves opportunity for harm to occur unchecked. Hash matching technology which is on the context of this consultation solely used for the removal of CSAM, should also be expanded to include other forms of illegal content most notably A.11 Adult image-based sexual offences, and non-priority offences. Tools such as StopNcii which is operated by SWGfL (A partner at the UK Safer Internet Centre) should be encouraged and used by smaller services with high or medium risk functions such as messaging, URL sharing and video-sharing platforms. While we appreciate that the burden on smaller platforms can be more significant, StopNCII.org provides support for smaller platforms with technical implementation from our internal web team, and potentially also from our larger existing partners. We plan to add different types of hash to the process to increase the number of platforms who can join and increase the protection to users. Multiple hash types also give greater protection to users by increasing accuracy and improving identification where some editing of images has occurred.
Question 15.1: Do you agree with our proposals? Please provide the underlying arguments and evidence that support your views.	No Comment

[50] Fuzzy matching can allow a match between U2U content and a URL list, despite the text not being exactly the same.

Question (Volume 4)	Your response
Question 16.1: Do you agree with our proposals? Please provide the underlying arguments and evidence that support your views.	No Comment
Question 17.1: Do you agree with our proposals? Please provide the underlying arguments and evidence that support your views.	*Is this answer confidential? No* <u>Blocking and default settings</u> In general, UKSIC agrees with the proposals however, we disagree with the "Enhanced User Control" provisions for small services. In particular ref. 9A and 9B which refer to "Users are able to block or mute other individual users and be able to be uncontactable by users they do not yet have an on-service connection with", and "Users can disable comments relating to their own posts, including comments from users that are not blocked". 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 something that Childnet has actively promoted, and is a tool they would use first, ahead of reporting for example, and we have shown this in our own research[21] that it preferrable over reporting, and the logic is that it provides a user control of a situation in a way that reporting doesn't. We would therefore recommend that blocking tools should be a requirement for all services including small services. Childnet's research[22] 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

Question (Volume 4)	Your response
	would report". This clearly showcases the importance of ensuring that blocking remains an option for children in all platforms including small services[23]". 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. 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 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.
Question 17.2: Do you have any evidence, in particular on the use of prompts, to guide further work in this area?	No

Question (Volume 4)	Your response
Question 18.1: Do you agree with our proposals? Please provide the underlying arguments and evidence that support your views.	*Is this answer confidential? No* Safety-by-design In general, however we disagree with the industry-centric approach of this consultation, instead we would prefer to witness and contribute to the adoption of 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[41], there is clear evidence that children are affected negatively when using such platforms: (36%) of children notice a negative change in their mental change 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. Recommend 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. We also urge 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. According to Schedule 4, Paragraph 13 (6), Ofcom is required to consider the accuracy, effectiveness, and impartiality of the technology employed. While Ofcom has effectively outlined how Hash Matching and URL blocking meet these criteria, it would also be beneficial for Ofcom to adopt a similar approach when recommending technologies for enforcement actions under Section 122 of the Bill, particularly in relation to the Use of Technology Notices, which are yet to be developed.

Question (Volume 4)	Your response
Question 18.2: Are there functionalities outside of the ones listed in our proposals, that should explicitly inform users around changing default settings?	*Is this answer confidential? No* Safety-by-design In general, however we disagree with the industry-centric approach of this consultation, instead we would prefer to witness and contribute to the adoption of 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[41], there is clear evidence that children are affected negatively when using such platforms: (36%) of children notice a negative change in their mental change 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. Recommend 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. We also urge 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. According to Schedule 4, Paragraph 13 (6), Ofcom is required to consider the accuracy, effectiveness, and impartiality of the technology employed. While Ofcom has effectively outlined how Hash Matching and URL blocking meet these criteria, it would also be beneficial for Ofcom to adopt a similar approach when recommending technologies for enforcement actions under Section 122 of the Bill, particularly in relation to the Use of Technology Notices, which are yet to be developed.

Question (Volume 4)	Your response
Question 18.3: Are there other points within the user journey where under 18s should be informed of the risk of illegal content?	*Is this answer confidential? No* Safety-by-design Targeted advertising, as highlighted in Ofcom's initial draft register, is identified as a potential enabler of various illegal harms covered by the legislation. It involves services gathering user data to create personalized profiles for delivering highly tailored advertisements. Some of these ads might promote products or services that are harmful to children's health and well-being, violating the Age-Appropriate Design Code (AADC), including those with age restrictions. To address this issue, advertising targeting children's accounts should be disabled, and there should be clear indications when content is sponsored or paid for.
Question 19.1: Do you agree with our proposals? Please provide the underlying arguments and evidence that support your views.	No comment
Question 19.2: What evaluation methods might be suitable for smaller services that do not have the capacity to perform on-platform testing?	No comment
Question 19.3: 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. Are you aware of any other design parameters and	*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

Question (Volume 4)	Your response
choices that are proven to improve user safety?	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]
Question 20.1: Do you agree with our proposals? Please provide the underlying arguments and evidence that support your views.	*Is this answer confidential? / No* <u>Alternative Dispute Resolution</u> 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 can be found in volume 3 above. 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[39] 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

Question (Volume 4)	Your response
	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.
Question 20.2: Do you think the first two proposed measures should include requirements for how these controls are made known to users?	No Comment
Question 20.3: Do you think there are situations where the labelling of accounts through voluntary verification schemes has particular value or risks?	*Is this answer confidential? / No* What could also be linked to sexual exploitation of children, is the proposed guidance for verification schemes. According to user Control Ref.9C[34] we disagree with the lack of requirements set out for smaller services for internal and transparent policies regarding their verification and paid-for-verification scheme. Implementing a verification scheme can be useful, especially with issues like misinformation which are mentioned in volume 2. However, the lack of a transparent process which is not a requirement for most services could lead to exacerbated danger for children and could amplify risk for issues such as grooming and sextortion (Understanding Verification on Instagram). According to the BBC Bitesize research[35], 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

Question (Volume 4)	Your response
	to impersonate official sources and deceive users. Specifically, reporting on X Verification has revealed vulnerabilities to scams within these schemes. Ofcom's research[36] indicates that nearly a quarter (23%) of 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.
Question 21.1: Do you agree with our proposals? Please provide the underlying arguments and evidence that support your views.	No Comment
Question 21.2: Do you have any supporting information and evidence to inform any recommendations we may make on blocking sharers of CSAM content? Specifically: • 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)? What are the advantages and disadvantages of the different options, including	*Is this answer confidential? / No* Due to the lack of uniformity and substantiated data on the current utilization of this strategy by services, we suggest that Ofcom leverages its authority to gather information. This would help ascertain how regulated services are employing methods like blocking or issuing strikes against users who violate laws or terms of service. Such insights can inform future updates to the Code. Additionally, this assessment should address obstacles like VPN usage and devise strategies to deter users from creating new profiles, particularly to counteract perpetrators employing burner accounts. Providing precise guidelines with clear criteria would assist services in implementing these measures proportionately and uniformly. CSAM

Question (Volume 4)	Your response
any potential impact on other users? • How long should a user be blocked for sharing known CSAM, and should the period vary depending on the nature of the offence committed? • 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. 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?	According to the research of IWF[17] and SWGfL[18] a lot of the services that host CSAM or Adult Intimate Abuse Images are hosted abroad, and are operating services in international legal loopholes, where the international policing and Inhope network do not have access to. How can this OSA escalate and assist the process of removal of such content that could even be self-generated by UK citizens. 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[19] published by the IWF illustrates that A.I poses a significant risk particularly with the potentially exacerbated volume of csam images 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[20], that should reflect a global online safety regime, where the risk and harm will be minimised. Smart tools and resources such as Stop-Remove and Stop Ncii, should be encouraged to tackle the exacerbated risks that evolving technologies pose on services and children.
Question 22.1: Do you agree with our proposals? Please provide the underlying arguments and evidence that support your views.	*No comment*
Question 23.1: Do you agree that the overall burden of our measures on low risk small and micro businesses is proportionate?	No comment

Question (Volume 4)	Your response
Question 23.2: 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?	No comment
Question 23.3: We are applying more measures to large services. Do you agree that the overall burden on large services proportionate?	*Is this answer confidential? / No* **The importance of Safety-by-design** As the UK Safer Internet Centre, safety online is at the core of our work. On the context of this consultation, the UKSIC would like to acknowledge that the scale of the work of navigating such a complex topic can be really difficult and we would like to acknowledge the efforts of the Ofcom team. Overall, however UKSIC would like to note the fact that the focus of this consultation is industry-centric, which does not reflect the harmful nature and above all the victims of illegal harms online. With that idea in mind, we instead propose a victim-centric approach which will provide a safety-by-design framework, that will facilitate the transition to a safer digital environment. The ESafety Commissioner [2] has created a series of principles which accompany the Safety-by-design process which the UKSIC would like to see the inclusion of. Although we comprehend the approach and classification of risk by size[3], 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. As noted in Volume 2, women and minorities are a lot more likely to face harm and provisions should be put in place to protect them from harm. **Children Illegal Harms** 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

Question (Volume 4)	Your response
	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.
Question 24.1: 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? If not, why not?	No Comment

Question (Volume 5)	Your response
Question 26.1: Do you agree with our proposals, including the detail of the drafting? What are the underlying arguments and evidence that inform your view.	<u>Alternative Dispute Resolution</u> Children affected by a service's design features should have avenues for recourse, considering their vulnerability to various online harms. Once a child encounters content or activities that breach a service's legal safety obligations under the Act, prompt reporting and resolution are imperative. However, the Act lacks provisions for individuals to lodge complaints with regulatory authorities or advocacy bodies when they've suffered harm. Existing reporting mechanisms are failing children, particularly: • Research by the Children's Commissioner[15] for England revealed that 40% of children refrained from reporting harmful content because they believed it would be futile. Additionally, 30% cited a

Question (Volume 5)	Your response
	lack of knowledge on how to report, while 25% were unaware that the content could be reported. Only 15% felt that reporting was unnecessary. • The same research found that platforms often overlook children's reports. Merely 63% of children reported that the content they flagged was removed, while 25% observed no action taken, and 10% were unsure of any outcomes resulting from their reports. This underscores the need for independent appeals as a component of the Online Safety Act. UKSIC would also like to share the concern raised by the SWGfL in relation to the recent report from the Public Accounts Committee[16], which highlighted that it could be years before the public saw any demonstrable change in their online lives. "Ofcom prepared well for its new responsibilities, and moved swiftly to implement the OSA when it became law in October 2023. But the PAC warns of potential public disappointment with the new regulatory regime, which will not be fully implemented until 2026, if people cannot quickly see improvements to their online experience or understand how complaints are acted on. With Ofcom able only to take action where there are systemic concerns about a service provider, the report recommends it develop a mechanism for letting people know what impact their complaint has had". Dame Meg Hillier MP, Chair of the Committee, said: "*Expectations are understandably high for firm guardrails in the hitherto largely unregulated online world. We know that around two thirds of UK children and adults say they experienced at least one potential online harm in a month in 2022, according to Page 11 Ofcom, which is to be commended for how swiftly it has moved to take on its new responsibilities. It must now continue to be proactively frank with the public over what the Online Safety Act does and does not empower it to do, lest confidence in the new regime be swiftly undermined.*" "Firm detail on how fees for industry, enforcement, automated monitoring and a range of other issues must now be locked in. No other country has introduced equivalent online safety regulation. Ofcom now needs to capitalise on its early progress. It must also accelerate its

Question (Volume 5)	Your response
	coordination with other regulators both at home and overseas, in the recognition that it is at the forefront of a truly global effort to strike the right balance between freedom and safety online."
Question 26.2: Do you consider the guidance to be sufficiently accessible, particularly for services with limited access to legal expertise?	No Comment
Question 26.3: What do you think of our assessment of what information is reasonably available and relevant to illegal content judgements?	No Comment

Question (Volume 6)	Your response
Question 28.1: Do you have any comments on our proposed approach to information gathering powers under the Act?	No Comment
Question 29.1: Do you have any comments on our draft Online Safety Enforcement Guidance?	*Is this answer confidential? / No* **Children Illegal Harms** It is encouraging that, in accordance with the law, the choice to pursue enforcement measures will focus on instances where the service has violated its obligations regarding child safety.

Question (Volume 6)	Your response
	Concerning paragraph 29.39(b), when assessing whether children can access certain parts or the entirety of the service, Ofcom must verify that the age verification methods implemented by the service adhere to the standards outlined in the age assurance guidance. Merely having age verification mechanisms in place does not guarantee that children cannot access the service; the effectiveness of these measures depends on their quality. The enforcement process should act as a mechanism that will protect children from accessing harmful and age-inappropriate content. Various international examples of exemplary practices exist regarding the integration of safety by design. For instance, the Australian e-safety commissioner[37] has formulated principles, an accessible assessment tool for services, resources tailored for investors and financial entities, and guidance for the tertiary sector on effectively engaging all relevant components of a safety-by-design process. We encourage Ofcom to contemplate a similar strategy for regulation in the UK. Although there is some indication of adherence to safety-by-design principles, such as the proposed Grooming mitigations in Ofcom's code of practice, these measures are presently limited to the largest platforms or those deemed to be at medium to high risk of Child Sexual Abuse Material (CSAM). We perceive it as a missed opportunity not to capitalize on the successes of implementing the Age-Appropriate Design Code[38] to ensure that platforms incorporate safety measures from the outset, rather than constantly having to retrofit solutions to combat the spread of illegal content on their platforms. <u>Alternative Dispute Resolution</u> UKSIC would also like to share the concern raised by the SWGfL in relation to the recent report from the Public Accounts Committee[16], which highlighted that it could be years before the public saw any demonstrable change in their online lives. “Ofcom prepared well for its new responsibilities, and moved swiftly to implement the OSA when it became law in October 2023. But the PAC warns of potential public disappointment with the new regulatory regime, which will not be fully implemented until 2026, if people cannot quickly see improvements to their online experience or understand how complaints are acted on. With Ofcom able only to take action where there are systemic concerns about a service provider, the report recommends it develop

Question (Volume 6)	Your response
	a mechanism for letting people know what impact their complaint has had". Dame Meg Hillier MP, Chair of the Committee, said: "*Expectations are understandably high for firm guardrails in the hitherto largely unregulated online world. We know that around two thirds of UK children and adults say they experienced at least one potential online harm in a month in 2022, according to Page 11 Ofcom, which is to be commended for how swiftly it has moved to take on its new responsibilities. It must now continue to be proactively frank with the public over what the Online Safety Act does and does not empower it to do, lest confidence in the new regime be swiftly undermined.*" "Firm detail on how fees for industry, enforcement, automated monitoring and a range of other issues must now be locked in. No other country has introduced equivalent online safety regulation. Ofcom now needs to capitalise on its early progress. It must also accelerate its coordination with other regulators both at home and overseas, in the recognition that it is at the forefront of a truly global effort to strike the right balance between freedom and safety online."

Question (Annex 13)	Your response
Question A13.1: 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?	No comment
Question A13.2: 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	No comment

Question (Annex 13)	Your response
and treating Welsh no less favourably than English.	

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