Your response

Please refer to the sub-questions or prompts in the annex to our call for evidence.

Question	Your response
Question 1: Please provide a description introducing your organisation, service or interest in Online Safety.	*Is this response confidential? – N* The NSPCC is UK's leading child protection charity with over 130 years in experience safeguarding children from harms. We have led the campaign for online regulation and were a driving force in the introduction of the Online Safety Bill. We are committed to ensuring that children are safer online and intend to contribute to creating a regulatory regime which incentivises online service providers to embed safety by design when creating new products. We are committed to using our

	knowledge and expertise to raises children's voices in the debate and support the development of a strong, child-centric, regulatory framework which reduces harm for children online.
Question 2: Can you provide any evidence relating to the presence or quantity of illegal content on user-to-user and search services? **IMPORTANT: Under this question, we are not seeking links to or copies/screenshots of content that is illegal to hold, such as child sexual abuse. Deliberately viewing such images may be a criminal offence and will be reported to the police.**	*Is this response confidential? – N* **<u>Child sexual exploitation and abuse</u>** The scale of online abuse and the presence of Child Sexual Abuse Material (CSAM) has grown significantly in recent years. Since 2017, the NSPCC has attempted to monitor and track the prevalence of child sexual exploitation and abuse (CSEA) offences. Through NSPCC Freedom of Information request to police forces in the UK and Home Office official statistics, we know that between 2017/18 and 2021/22: • There has been a 127% increase in online child sexual abuse crimes. • There has been an 84% rise of online grooming crimes in the UK. • Girls represented at least[1] 82% of grooming cases in 2021/22 with 39% of these victims being between 12-15 yr. old. Youngest recorded age of online grooming in 2021/22 was 4 yr. old. • There has been a total of 107,555 reported cases of indecent images in the UK. We know that abusers frequently take advantage of the system and design of user-to-user online service providers to commit their abuse. The NSPCC tries to collect platform level data that is available to identify 'high-risk' online services. Our internal findings show that in 2021/22 alone police forces recorded 70 different apps and games being involved in online grooming offences. Police forces also cite the following online service as being most frequently used by perpetrators when committing abuse[2]. • Snapchat: 33% • Instagram[3]: 20% • Facebook and Messenger[3]: 9% • WhatsApp[3]: 9% • TikTok: 6% • Kik: 4.5% • Discord: 3% From speaking with children, victims, and survivors. We know that abuse is not siloed to one particular platform and often happens across many in parallel. We refer to this as cross-platform risk.

Childline heard of offenders seeking to redirect conversations from a public online space, such as a forum or group chat, to a private online space, including end-to-end encrypted channels and private live streaming – a common tactic groomers use to avoid detection. Below is an excerpt from a 15-year-old girl who contacted Childline for advice:

"We met around the start of quarantine on an online video game. It was purely by coincidence that we ended up on the same team. We did quite well so we decided to party up for another round. There is an in-game voice chat function for those in a party, so we ended up being friends over the next few games. We had a lot of conversations via chat, and after a few months we transitioned to Discord then eventually WhatsApp. We now talk every day and have made plans to meet in person. I need advice about it from others as most people do when dating someone, but no one seems to be able to look past his age."

The detrimental impact of limited industry regulation cannot be overstressed. Through NSPCC's helpline and counselling service, Childline, we have been at the frontlines of supporting children to protect themselves from the harms of the online space. Childline offers us a unique opportunity to learn about the emerging harms and the child-specific impacts. For instance, we found that around 50% of the calls children made regarding online-related issues in 2021/22 were about CSEA concerns.

Some of the key CSEA concerns our counsellors recorded were about:

- Blackmail/threats to expose/share sexual images
- Sexting/sharing self-generated sexual images
- Grooming/sexual exploitation online
- Received nudes/explicit images

Where Childline counsellors heard references to CSAM, these typically occurred in the context of young people receiving unsolicited indecent images of children online, or where young people had come across such content themselves unintentionally (in the course of browsing online service providers which they believed to be safe).

We are also seeing more evidence of contextual CSA, otherwise referred to as 'CSA breadcrumbing'.[4] CSA breadcrumbing is content that directly facilitates CSA but is not illegal itself. It a type of activity where abusers form offender networks by posting 'digital breadcrumbs' that signpost their interest in children and to illegal child sexual abuse content elsewhere online. This was brought into the scope of the Bill via Government led amendments 58, 59, 60, 61 and 102 during report stage of the House of Commons.

CSA breadcrumbing includes techniques such as:

- Tribute sites: where abusers create social media profiles using misappropriated identities of known child abuse survivors. These are used by offenders to connect with like-minded perpetrators, to exchange contact information, form offender networks and signpost to child abuse material elsewhere online. Internal report1 found that in the first quarter of 2021, there were 6 million interactions with such accounts.

- Signposting abuse on social networks: abusers are increasingly using novel forms of technology to signpost to abuse, including QR codes.

Other illegal offences

Other illegal offences recorded by our counsellors include pro suicide content, threats of violence and death, cyberstalking. Listed below are excerpts from Childline counselling sessions:

Pro-suicide content
Childline counsellors heard from young people who had encountered links to dangerous online challenges in which teenagers are encouraged to perform a series of increasingly extreme tasks including self-harm and even suicide.

"Me and my friends keep getting added to these weird accounts on TikTok. Basically, they keep asking us to become involved with a challenge and if we say no, they threaten to hurt us and our family. I've tried blocking them, but then more accounts start popping up again. What should we do about it because they said they're gonna kill me in my sleep!" (Girl, 13)

Some young people were worried that their failure to participate in these challenges would lead to further malicious advances online; some shared a fear of being hacked, traced, or physically threatened by the people behind the challenge.

Threats of violence and death
Childline heard from young people who had been subjected to threatening and intimidating comments on social media; in some cases, young people spoke about receiving threats of violence, death threats and/or messages telling them to kill themselves.

"I used to be friends with this guy on Instagram who I don't know IRL. We were kind of dating until he started acting weird, so I blocked him. Then he sent me these DMs saying he wanted to kill me and all my family! I don't know if he's bluffing, but he said he

knows where I go to school – I don't know how since I never talked about that stuff. I'm so scared for my life, I don't want to leave the house!" (Girl, 14)

Cyber-stalking

Some young people who contacted Childline believed they were being cyber-stalked, whether by someone they knew or a stranger online, often across multiple online service providers. Some young people were frightened at the level of personal information their stalkers seemed to know about them. In some cases, young people received threatening messages.

"I'm being stalked online by a stranger who seems to know everything about me in detail. They know my address, where I go to school and what I look like. They say that they are watching me. The person says very sexual things and crazy stuff. I am so scared I can't breathe properly, am shaking and can't sleep or eat. It is making me paranoid and I am terrified about leaving the house on my own." (Girl, 14)

Information gathering

The insights of the NSPCC are from a combination of Childline and the NSPCC helpline, research, freedom of information requests, and anecdotal evidence from trusted sources. This is, therefore, only likely to be the tip of the iceberg. Thorough analysis of such systemic harms can only be made when platform level data is shared by service providers themselves. While the NSPCC has some insights which helps it understand the landscape, there are limitations to this data and there is much still to learn.

For instance, current police record keeping systems do not capture the full scale of cross-platform abuse. Despite abuse often happening on multiple online service providers, police records will often only note the platform the offence started on. This both misrepresents the scale at which abuse is happening and does not appropriately capture the probability of perpetrators' conducting cross-platform abuse.

Greater transparency of data from service providers and analysis of this data by civil society is needed to build a true understanding the scale, type, and patterns of harm online (see Q27 for more detail).

Recommendations

To improve the outcomes to children, we need a transparent regulatory regime where civil society can hold companies to account. We recommend that Ofcom ensures:

- Publication of risk assessments: Civil society and academics should be able to access regular reports of

	harms identified by each platform and the appropriate interventions/responses they have made. • Standardisation of reporting: We would also encourage standardised reporting measures to ensure there is a consistency in reporting. Footnotes: 1 Results are based on findings where gender was known and/or recorded. 2 Results based on findings where the online service provider is known. 3 Meta-owned online service providers. 4 CSA breadcrumbing was brought in scope of the Online Safety Bill through amendments 58, 59, 60, and 61 (amendments relating to the 'commission or facilitation of an offence' and service design).
Question 3: How do you currently assess the risk of harm to individuals in the UK from illegal content presented by your service?	*(Not applicable)*
Question 4: What are your governance, accountability and decision-making structures for user and platform safety?	*(Not applicable)*
Question 5: What can providers of online services do to enhance the clarity and accessibility of terms of service and public policy statements?	*Is this response confidential? – N* While this question is intended for service providers, we recommend the following steps to enhance clarity of terms of services (ToS) and public policy statements: **Child-friendly terms of service** The ICO has guidance[5] on how information related to GDPR should be presented to children. These same principles could be applied to Terms of Service. Examples include: • Having different versions of the policy for different age groups, or ensuring that the youngest age group can understand the messages • Suggests making it interesting to children to read, such as using diagrams, pictures and videos **Accessibility of information** The accessibility of this information is also important – it should be readily available and easy to refer back to at all times, rather than having to be sought out. Guidance should consider how children with different needs may need the information presented in different ways and should take steps to make this

	available to and understandable by all children. For example, basic accessibility standards should be met, such as ensuring they can be read by a screen reader. Terms of service should also be run through software that assesses the reading age. The reading age should be no higher than the possible youngest user. This would support young people who may be older but have additional needs such as a lower reading age. Simplified social media terms and conditions should be considered as basic requirements. The Children's Commissioner (2017) report[6] can be used as a reference model. **Recommendation** • Accessible and child-friendly statements: To ensure children are safe online, terms of services and public policy statements must be clearer. Footnotes: 5 ICO (n.d.) How does the right to be informed apply to children? 6 Children's Commissioner (2017) Simplified social media terms and conditions for Facebook, Instagram, Snapchat, YouTube and WhatsApp.
Question 6: How do your terms of service or public policy statements treat illegal content? How are these terms of service maintained and how much resource is dedicated to this?	*(Not applicable)*
Question 7: What can providers of online services do to enhance the transparency, accessibility, ease of use and users' awareness of their reporting and complaints mechanisms?	*Is this response confidential? – N* Although reports and complaints can form useful insights, evidence[7] suggests that most children do not use reporting or complaints mechanisms when they have been subjected to potentially harmful experiences online and, therefore, we should not wholly be reliant on these mechanisms for safeguarding children. Instead, online service providers need to embed a range of solutions need to protect children. We know that **online sexual abuse is often not disclosed but is instead discovered**. Children are often unaware of the dynamics of being groomed and are therefore it is doubtful that they will report or complain. This means that the most significant online harms are unlikely to be captured by reporting or complaints mechanisms. As we outline in our User Advocacy (2022)[8] report, a survivor of abuse says "[w]e need to stop putting the responsible on a vulnerable child to prevent crime" (p.8). This means that

other tools, such as user advocacy are needed to ensure that survivors experiences inform product and regulatory decisions.

The Thorn report (2021) found that primary reasons children do not use reporting functions is because they either (a) do not consider the offence as an issue, (b) are worried about whether they will remain anonymous, or (c) are embarrassed at having experienced the offence[7]. Ofcom (2022) findings[9] indicate that only a third of children knew how to use online reporting or flagging functions (32%); and just 14% had ever used them. The report also found that nearly all children aged 12-17 were aware of at least one safety feature to help keep themselves safe online (94%); 84% had put these into practice. Blocking people on social media was the behaviour with the highest levels of awareness and use. According to the Thorn report (2021), children saw blocking as a form of self-protection whereas reporting was seen as a form of punishment for actions that broke the rules.

Evidence[10] also suggests that gender norms play a role in this issue with girls regularly being shamed and victim blamed for sharing image and that the fear of this often prevents reporting. There are also fears among young people that reporting may lead to them being removed off of the platform or making matters worse by taking away their agency to proceed how they want to. Below is an excerpt from a Childline counselling session where a 16-year-old girl voices her doubts about reporting:

"I feel I have been peer pressured to share inappropriate images of myself online. I'm now being blackmailed with images that I've sent in the past. I was 15 years old when the images were taken. I've considered going to the police, but I'm unsure of the process and whether it's worth it, now that I'm 16."

Online service providers should make a concerted effort to understand the dynamics of abuse and why children are not using the reporting and complaints mechanisms. This information should then be translated into redesigning their complaints tools to ensure reporting is more accessible[11].

Recommendations

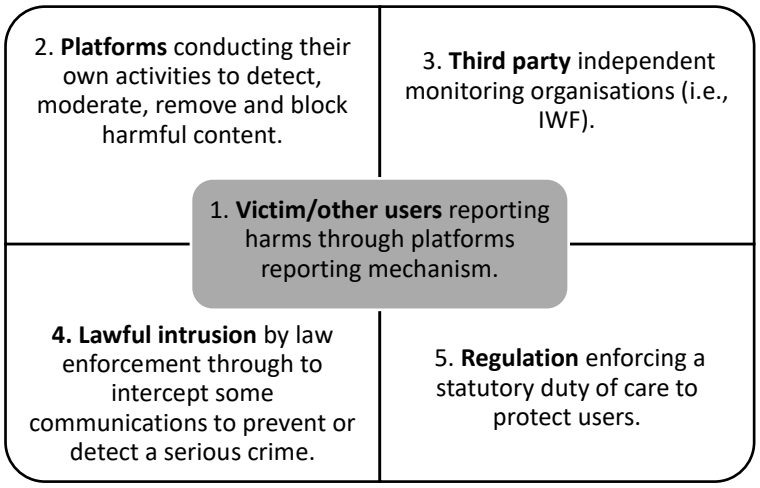
There needs to be multiple tools for ensuring that children are safe online and not placing the responsibility on children to prevent harm. These include:

- User advocacy: Strong user advocacy arrangements for children will help ensure that children views inform the

	future regulatory decisions, and the regime delivers positive outcomes and reduces harm for children. • Multiple methods of reporting: Online service providers should be offering multiple accessible reporting options to children. Online service providers should also be doing more to highlight different options such as muting which could play a vital role in young people's friendships online. • Cultural change in reporting harms: Online service providers have a responsibility to tackle the normalisation of online harms. Reassurances over anonymity and confidentiality should be offered as much as possible. Online service providers need to take a victim-centred approach to dealing with reporting and complaints mechanism. Footnotes: 7 Thorn (2021) Responding to Online Threats: Minors' Perspectives on Disclosing, Reporting, and Blocking. 8 NSPCC (2022) Making the case for user advocacy: NSPCC's proposals for user advocacy arrangements in the Online Safety Bill. London: NSPCC. 9 Ofcom (2022) Children and parents: media use and attitudes report 2022. London: Ofcom. 10 Ringrose et al (2021) Understanding and Combatting Youth Experiences of Image-Based Sexual Harassment and Abuse. London: UCL Institute of Education. 11 Creating visually appealing and easy to follow way of reporting is an effective way to make content accessible. Keep Cup's environmental impact calculator is a good example of how this can be achieved.
Question 8: If your service has *reporting or flagging* mechanisms in place for illegal content, or users who post illegal content, how are these processes designed and maintained?	*(Not applicable)*
Question 9: If your service has a *complaints* mechanism in place, how are these processes designed and maintained?	*(Not applicable)*

Question 10: What action does your service take in response to *reports* or *complaints*?	*(Not applicable)*
Question 11: Could improvements be made to content moderation to deliver greater protection for users, without unduly restricting user activity? If so, what?	*Is this response confidential? – N* Moderation is one of the most effective tools for protecting children because it both removes harmful content and protects children from abuse. Improvements to both human and automation content moderation is needed to ensure they remain an effective safeguarding tool: **Human moderation** There are potential pitfalls to human moderation. Some online service providers rely primarily on volunteers in their own communities to self-police, with administrators doing occasional checks. There is a real concern here that the subjective nature of this moderation process creates inconsistency and potential for gaps in protection of users. Reddit is an example of this and demonstrate reluctance for tougher moderation as it contravenes their philosophy which is 'community over content.' There are content rules enforced by administrators but decisions about content removal lie with the community moderators and so there is potential for inconsistency in what is considered acceptable and room for interpretation of the content rules. For volunteer or professional human moderators to be effective, they should receive training so they can discharge their duties effectively and consistently. They should also receive training specific to the content they are moderating. For instance, moderators looking at CSA content and activities should be trained in moderation and safeguarding. **Improvements in automation** Automated moderation can be used in combination with human moderation to improve how platforms detect and takedown harmful content through the internal data online service providers have available on their services. There are current projects, like Project Artemis and co-NSPCC project Dragon-S, which explore how grooming behaviours can be moderated. Both programmes use AI and linguistics to understand patterns of grooming behaviour. This has the potential to flag risks to law enforcement or child protection professionals and could also be used to flag to content moderators. Such tools could also be expanded to look for illegal or harmful content.

	Getting the balance right Effective safeguarding requires a balance between human moderation and automation that online service providers should constantly tweak dependant on the risk levels on their platforms. **Collaboration** Collaborations with external moderators such as trusted flaggers and third party moderated should also be used to ensure a matrix of moderation tools are available to deliver greater protection. For instance, the NSPCC operate a Trusted Flagger process where we have a direct link with some of the online service providers to prioritise content that should be removed. This is generally used when members of the public share content directly with either of our counselling services, Helpline and Childline, to request take down of content or voice concern about the content. However, we have recently found issues with this process where an undue level of information is requested from the online service providers, or content has not been removed, even though a trusted body has said this is causing harm. It is essential that online service providers ensure reports from Trusted Flaggers operate effectively by responding swiftly to content that has been flagged without demanding undue level of information. We are conscious that groups will use freedom of as an argument to restrict moderation speech by claiming that the greater the burden for content moderation, the greater the risk to freedom of speech. **However, we think this approach is incorrect as it ultimately pits safety against freedom of speech, when such a binary does not have to exist.** Online service providers have both the internal skills and resources, and external support from expert bodies, to create solutions which appropriately balance both. **Recommendations** • Ensure trusted flagger programmes are effectively streamlined • Publication of risk assessments: (see Q2).
Question 12: What automated moderation systems do you have in place around illegal content?	*(Not applicable)*

Question 13: How do you use human moderators to identify and assess illegal content?	*(Not applicable)*
Question 14: How are sanctions or restrictions around access (including to both the service and to particular content) applied by providers of online services?	*Is this response confidential? – N* The following response is based on the regulators request to receive evidence around the safeguards we consider should be in place to protect users' privacy[12]. **Safeguarding limitation posed by end-to-end encryption** We are conscious that a number of respondents will highlight end to end encryptions ability to safeguard users privacy. However, while encryption measures can generally be seen as a protective layer for sharing sensitive information online, the growing adoption of end to-end encryption (E2EE) by social media platforms can have profound negative impacts on online service providers ability to monitor and report child abuse and therefore safeguarding children. The existing approach to protecting children from online abuse replies on multiple elements outlined in the graphic below:  However, E2EE prevents or restricts the majority of existing lines of defence to online harms. A truly end-to-end encrypted communication would only be accessible by the device (and therefore the person) sending and the device (and person) receiving the message; neither the hosting platform nor law enforcement can see its content. Present online safety framework will be restricted with E2EE if the content of the messages cannot be moderated. Risks associated with E2EE can be mitigated and we should not be led into a false binary between children's safety and privacy[13]. The recent GCHQ report[14] by Ian Levy and Crispin Robinson discusses how CSAM could be detected within encrypted

	services whilst maintaining user privacy, including the potential use of client-side image scanning. **Discrepancy in digital access for children with Special Educational Needs or Disability (SEND)** There are a number of legitimate circumstances where online services will restrict a user form accessing content. For instance, to restrict children from age-inappropriate content. However, depending on the technology used to assure/verify the age, it can have knock on implications and has the potential to restrict SEND children from accessing the services. Platforms need to be conscious of these dynamics and build mitigations accordingly so that groups are not wrongfully restricted. Footnotes: 12 Ofcom (2022). Call for evidence: First phase of online safety regulation. 13 For further details on risk mitigations see NSPCC (2021) End-To-End Encryption: Understanding the impacts for child safety online. London: NSPCC. 14 Levy, I. and Robinson, C. (2022) Thoughts on Child Safety on Commodity Platforms. doi.org/10.48550/arXiv.2207.09506.
Question 15: In what instances is illegal content removed from your service?	*(Not applicable)*
Question 16: Do you use other tools to reduce the visibility and impact of illegal content?	*(Not applicable)*
Question 17: What other sanctions or disincentives do you employ against users who post illegal content?	*(Not applicable)*
Question 18: Are there any functionalities or design features which evidence suggests can effectively prevent harm, and could or should be deployed more widely by industry?	*Is this response confidential? – N* **Safety-by-design as an industry wide approach** A cultural overhaul is required where safety-by-design is made a key principle when online services consider updating or creating new design features and functionalities. Currently, design features and functionalities are not made with safety and harm mitigations in mind, instead the matrix of solutions (see Q14) to online safety are often retrofitted within profit-orientated ecosystem. However, from design to implementation it wholly feasible to build safe online services. Should industry choose not to start from a principle of safety by design, we run the risk of creating an online world in which **children themselves are the main line of defence against their own abuse** and we rely on victim reporting alone which is already limited (see Q7). For us to

adequately safeguard children, safety must become a priority for all online services. To embed a safety-by-design approach all service providers should ensure:

1. **Risk assessments include third party input, and the final output is made publicly available.** Online service providers should be expected to risk assess how high-risk features operate on their services and demonstrate that the functionality is safe for children to use. If a platform cannot demonstrate that appropriate risk mitigations are in place, it should consider whether it is appropriate to continue offering it.

   For example, recent updates to Facebook Messenger now include end-to-end encryptions (E2EE) on chats. With Facebook Messenger being used in a significant number of online CSEA offences (see Q2), this new design feature will further place children at risk when interacting in this space (see Q14 for risks of E2EE). This is a clear example of where risk assessments with input from child protection charities would lead to better outcomes for children.

2. Appropriate safeguarding mitigations are included for proven high-risk design features such as livestreaming, private messaging, and E2EE.

   The regulator should maintain a list of high-risk design features and update this regularly. Due to the visual and inherently unpredictable nature of livestreaming services, children are at risks of CSEA on these platforms. With Snapchat representing over a third of online grooming offences (see Q2) in the UK it is indicative of the high risks livestreaming and video-chat services present high risks to children.

3. Age assurance and other minimum safeguarding practices are implemented consistently. Online service providers should utilise age assurance technologies to identify children so their accounts can receive the protections outlined above. Other minimum safeguarding measures already discussed in this form (see Q7) include:
   - Default privacy and safety settings for children's accounts;
   - Accessible, age-appropriate explanations of terms and conditions;
   - A transparent and responsive complaints process;
   - A dedicated reporting flow for complaints that relate to child abuse.

Question 19: To what extent does your service encompass functionalities or features designed to mitigate the risk or impact of harm from illegal content?	*(Not applicable)*
Question 20: How do you support the safety and wellbeing of your users as regards illegal content?	*(Not applicable)*
Question 21: How do you mitigate any risks posed by the design of algorithms that support the function of your service (e.g. search engines, or social and content recommender systems), with reference to illegal content specifically?	*(Not applicable)*
Question 22: What age assurance and age verification technologies are available to online service providers, and what is the impact and cost of using them?	*(Not applicable)*
Question 23: Can you identify factors which might indicate that a service is likely to attract child users?	*Is this response confidential? – N* In line with the safety-by-design ethos (Q18) **the default assumption should be that children may attempt to access any service**. The assumption that children are not using a particular service can only be made by providers if there is evidence that 100% of the users are not children. The regulator should not create a situation whereby a platform can claim exemption from their child safety duties because their platform was not intended to be access by children (i.e., OnlyFans).
Question 24: Does your service use any age assurance or age verification tools or related technologies to verify or estimate the age of users?	*(Not applicable)*
Question 25: If it is not possible for children to access your service, or a part of it, how do you ensure this?	*(Not applicable)*

Question 26: What information do you have about the age of your users?	*(Not applicable)*
Question 27: For purposes of transparency, what type of information is useful/not useful? Why?	*Is this response confidential? – N* **Publicly available data** The NSPCC and other civil society groups have been at the forefront of raising concerns about online harms through imperfect sources such as internal data streams, counselling helplines, and work with victims. In order to better understand the landscape for child protection, and hold companies to account, civil society must be able to access the data collected by online service providers on harms and offences identified. We must also be provided with risk assessments to ensure that online service providers are identifying and tackling known and reasonably foreseeable harms. With appropriate transparency, civil society can support online service providers in designing services that mitigate online harms. **User advocacy arrangements for children** Strong user advocacy arrangements for children[15] will help ensure that the future regulatory regime delivers positive outcomes and reduces harm for children. We believe that advocacy arrangements will ensure that: • Children's voices are heard and there is a funded mechanism for children to be able to channel their views • Safeguarding is front and centre of the new regulatory regime; • New and emerging risks to children are quickly discovered and tackled by the regulated companies and the regulator; and • There is an effective counterbalance to the technology sectors attempts to influence the regulations. **Recommendation** • User advocacy arrangements for children: user advocacy arrangements are needed to ensure that children's interests and experiences shape the regulatory regime. <u>Footnotes:</u> 15 NSPCC (2022) Making the case for user advocacy: NSPCC's proposals for user advocacy arrangements in the Online Safety Bill. London: NSPCC.

Question 28: Other than those in this document, are you aware of other measures available for mitigating risk and harm from illegal content?	*(Not applicable)*