May 2026

Growing up in the online world: Collective recommendations for tackling the risks of online child sexual abuse and exploitation

With the rapid development of technology, safeguarding in the digital age presents increased challenges. In the case of child sexual exploitation and abuse (CSEA) the digital environment, without appropriate guardrails, can impact the lives of children and families in the most horrific ways.

The consultation puts forward a series of options with regard to age-gating particular features and functionalities that could be harmful to children. While we recognise this early response is in efforts to disrupt the current scale of the challenge, Government must also remember that age gates are one component in tackling the ecosystem of online harms to children. On their own, no one intervention can tackle the challenge we face.

CSEA is a complex threat which is influenced by a range of different factors. The Centre of expertise on child sexual abuse (CSA Centre) explains that "children can be sexually abused in many different ways, by different people and in different places and situations, including online."[1] Factors can include, but are not limited to, the relationship between the perpetrator and the victim, characteristics that can make some children more vulnerable than others, the nature of the abuse, process involved in maintaining access to the victim, and where and how the abuse took place.[2&3] There is also the added consideration that perpetration can happen by anyone. The NSPCC states that "child sexual abuse is committed by men, women, teenagers and other children. Offenders come from all parts of society and all backgrounds."[4] This means that risk of CSEA cannot be pinpointed to any one single factor, such as harmful design choices. Rather, in the case of online CSEA, high risk features and functionalities can play a part in enabling and/or exacerbating the likelihood of harm.

[1] CSA Centre (2025). What you need to know about child sexual abuse [pdf]. p.5.
[2] CSA Centre (2020). A new typology of child sexual abuse offending [pdf].
[3] The National Archives (2022). The Report of the Independent Inquiry into Child Sexual Abuse [pdf].
[4] NSPCC Learning (2026). Protecting children from sexual abuse [web].

No single solution alone will be able to tackle the complexity, severity, and scale of online CSEA. At the start of this year, the NCA offered a sobering warning on the challenge ahead:

> "The scale and prevalence of the CSA threat has increased in severity and complexity over the years. It can occur in any community and across all social backgrounds. It remains one of the most significant threats across the UK."[5]

To turn the tide against these harms, Government must introduce a package of interventions which target different stages of the CSEA pathway: from prevention to tackling revictimisation.

The threat of illegal harms continues to increase in scale and severity. In 2025, IWF Analysts actioned 311, 610 reports which confirmed to contain or lead to CSAM.[6] The IWF has also seen a 260-fold increase in AI generated child sexual abuse videos with a 29% increase in category A videos – content depicting the most severe form of abuse.[7]

However, we are not starting from scratch in tackling the threat of online CSEA. The Online Safety Act (2023) still has enormous potential, but confidence in its ability to deliver meaningful change must be strengthened. The introduction of the first iteration of Ofcom's Illegal Harms Codes in March 2025 and Protection of Children Codes in July 2025 established a needed floor to begin addressing the risks to children online. These hard-won regulatory levers have been the product of years of campaigning from civil society organisations, like ours, who have been at the forefront of demanding better, faster, and more ambitious safeguards for children online. The next stage of strengthening children's safety online must fortify existing regulatory levers, strengthen legislative gaps, and introduce technical measures to achieve our shared goal of ensuring children's lives online are as safe as they are offline.

The aim of the consultation is to consider solutions for children that "add to, rather than take away, from their childhood."[8] To help Government achieve this ambition, the UK's leading civil society organisations dedicated to tackling online child sexual abuse and exploitation have prepared a list of priority technical, legislative and regulatory ('legislative') interventions.

Whilst this is a non-exhaustive list of interventions, they are urgently needed to tackle the complex threat of online CSEA. Many of us have already shared our detailed views in response to the Government consultation. We want to use this opportunity to further spotlight some of the measures that are a collective priority. Through these proposed interventions, Government has the opportunity to deliver transformational outcomes for children's safety online.

These recommendations consist of the following interventions:

Technical interventions:

- Introducing safeguards for end-to-end encrypted environments
- Mandating deterrence messaging across platforms

[5] NCA (2026). Child sexual abuse is increasing in severity, complexity and accessibility, say policing leads [news].
[6] IWF (2026). Internet Watch Foundation Annual Report 2025: Executive Summary [web].
[7] Ibid.
[8] DSIT (2026). Growing up in the online world: a national conversation. p.11.

- Introducing device-level nudity detection and blocking on children's devices
- Considering mandatory CSAM detection and blocking on all devices
- Adopting a risk-based approach to restrictions on features and functionalities

Legislative interventions:

- Regulating AI chatbots and companion services
- Establishing enforceable minimum age requirements
- Introducing a legislative framework for principle-based age-appropriate experiences
- Ensuring all aspects of products are safe by design
- Stopping unsafe products from reaching the market by strengthening the Online Safety Act
- Introducing a statutory stay-down provision in the Online Safety Act
- Amending Schedule 11 of the Online Safety Act to ensure all high-risk platforms, regardless of their size, are in scope of existing illegal content duties
- Setting minimum standards for the Terms of Service for Category 1 services
- Including an overarching duty of care in the Online Safety Act
- Placing a duty on Ofcom to cooperate with domestic regulators and placing the Digital Regulation Cooperation Forum on statutory footing

Collective recommendations from organisations tackling online CSEA

Together, the interventions below have the potential to provide life-altering outcomes for children's safety online.

In this section, we describe how our proposed interventions work, and in many cases, how they are already being applied within the digital environment. We also provide a brief analysis of the impact of each intervention against the CSEA harm types in the appendix.

Technical interventions

1. Introducing safeguards for end-to-end encrypted environments

End-to-end encrypted (E2EE) environments remain a high-risk functionality where inadequate safeguards and limited detection foster a safe haven for offending. The 2026 Protect Children report revealed that E2EE messaging services are some of the most highly used environments by offenders to access CSAM.[9] It is also widely understood that E2EE messaging sites are commonly used to groom children after initial contact has been made in public spaces.

> *"I'm in a serious situation that I want to get out of. I've been chatting with this guy online who's like twice my age. This all started on Instagram but lately all our chats have been on WhatsApp. He seemed really nice to begin with, but then he started making me do these things to 'prove my trust' to him, like doing video chats with my chest exposed. Every time I did these things for him, he would ask for more, and I felt like it was too late*

[9] Protect Children (2026). CSAM Perpetrator Research Report: Findings from a Survey of CSAM Perpetrators on Digital Platform Use and Design [pdf].

to back out… This whole thing has been slowly destroying me and I've been having thoughts of hurting myself." - Girl, aged 15, Childline[10]

While the risk of harms is clear, the ambition to tackle the CSEA threats in E2EE environments remains low. In the consultation Government note that E2EE services are not exempt from potential restrictions. We welcome further clarity on how age restrictions would be implemented within these environments. Particularly as there is a risk that preventing children from accessing online platforms, where age gates are easily introduced, could push more children into these environments.

Without tackling the existing and live threats in E2EE environments the ambitions of this consultation will be defunct. As a start, E2EE services must be mandated to detect and block CSAM before it can be encrypted. Pre-encryption checks, such as using an upload prevention method, provide a privacy-preserving way of detecting images and videos of child sexual abuse. Upload prevention is a technology agnostic method which is already adopted within encrypted environments. The IWF's 2025 explainer *Preventing the upload of child sexual abuse material (CSAM) in E2EE environments* outlines how services are using upload prevention to protect users from receiving malware and improve their user experience. [11] The explainer specifically highlights how hash-matching can be used as part of the upload prevention method. However, upload prevention method is not limited to hash-matching alone. It is long overdue for companies to extend their use of an upload prevention method to also detect for CSAM, as upload prevention is already being used to detect for other types of content.

Many major technology companies have the inhouse capabilities to detect and block novel images of child sexual abuse within encrypted environments. Meta (Instagram), Apple, and Google have all introduced nudity detection on to parts of their service.[12,13, and 14] For all three companies, the nudity detection technologies apply within selected encrypted environments but not all: Instagram Direct Messages, Apple iMessages and FaceTime, and Google Messages.[15] Despite these detection technologies not being applied across all of the companies E2EE services it is evident that there is already technical capabilities to do so.

Tackling grooming in E2EE requires a separate approach given that it is text-based and contextual. The NSPCC's 2025 report *Tools to combat online harms: protecting children in private messaging*[16] provides interventions across each stage of the grooming lifecycle to help

[10] Please note that Childline snapshots are based on real Childline service users but are not necessarily direct quotes. All names and potentially identifying details have been changed to protect the identity of the child or young person involved. This applies to all Childline snapshots used in this response.
[11] IWF (2025). Preventing the upload of child sexual abuse material (CSAM) in E2EE environments [pdf].
[12] Instagram (2024). New Tools to Help Protect Against Sextortion and Intimate Image Abuse [web].
[13] Apple (n.d.). Expanded Protections for Children [web].
[14] Google (2025). 5 new protections on Google Messages to help keep you safe [blog].
[15] Instagram direct messages (DMs) stopped providing the option to opt in for end-to-end encrypted communication as of May 2026. The nudity detection tool remained in use where chats where end-to-end encrypted.
[16] NSPCC (2025). Tools to combat online harms: protecting children in private messaging [pdf].

reduce the safety risks to children in encrypted and private messaging environments. We urge services to implement these grooming interventions.

Platforms operating in encrypted environments should not be exempt from foundational detection duties. We firmly believe the primary barrier to implementation of safeguards in encrypted environments is will, not technical capability. With clear evidence on the technical feasibility of implementing privacy preserving interventions in encrypted environments, Government must not delay the swift introduction of these obligations on all encrypted services.

2. Mandating deterrence messaging across platforms

Too many existing and proposed interventions place the burden of protection on children themselves. Platforms must also be required to implement preventative measures such as deterrence messaging which explicitly target users at risk of, or seeking to cause, harm. These warnings should be triggered at the moment of risk across all stages of the offender pathway. By directing deterrence messaging, nudges, warnings, and signposting at users who show concerning behaviour we target the intervention at the root of the problem - those with intent to cause harm. For example, the Lucy Faithfull Foundation's StopItNow service provides an anonymous helpline, email and chat services for anyone with concerns about child sexual abuse or people seeking information on how to prevent it.[17] This has led to positive behaviour changes amongst people who are at risk of offending or have displayed such behaviour.[18 & 19]

> *"Over the years I have been in and out of depression and recently suicidal. I used adult sites as a way to pass time and quiet the mind. I had a pop-up [warning] on an adult website... I had a look around at what you did and read some of the modules. About two months ago I gave up those sites and decided I want to keep my mind occupied and more productive. I found the modules on addiction and pornography very helpful. And since I am free from it I feel better in myself." – Ben,[20] Lucy Faithfull Foundation service user [21]*

Deterrence messaging and signposting are an important remedy in helping tackle risky or harmful behaviour at source. Ofcom's current approach to deterrence messaging is limited to requirements from only large search services.[22] However, this duty should be extended across all regulated platforms including encrypted environments.

> *"There is a misconception that encrypted services cannot address harm. In reality, we take meaningful action, including delivering deterrence messaging via Project Intercept,*

[17] Lucy Faithfull Foundation (n.d.). Helpline [web].
[18] Walsh, M., Denis, D., and Findlater, D. (2023). Deterring online child sexual abuse and exploitation: lessons from seven years of campaigning [pdf]. *The Faithfull Papers*.
[19] Lucy Faithfull Foundation (2026). Project Intercept Impact Report [pdf].
[20] Pseudonym
[21] Lucy Faithfull Foundation (2026). Project Intercept Impact Report [pdf]. p. 11.
[22] Ofcom (2025). Illegal content Codes of Practice for search services [pdf].

a science-backed approach we value." André Meister, Chief Technology Officer (NZ), Mega[23]

The success of deterrence messaging on platforms voluntarily deploying them provides a clear signal of the power these interventions have in preventing online child sexual abuse. To fully actualise their impact, Government should also provide further investment to deliver these interventions at scale.

3. Introducing on-device nudity blocking and CSAM detection

The 2025 report by the National Centre for Violence Against Women and Girls and Public Protection (NCVPP) reveals that indecent imagery of children (IIOC) offences is the most prevalent CSEA harm type.[24]

Introducing age-based restrictions at the platform level could tackle some aspects of the CSEA risks online by preventing children from being able to use their accounts to share nude images of themselves. For example, this would work through the use of age assurance methods that would allow a platform to determine whether the user was under or over 18. Where the platform determines the user is under 18, it would then trigger the needed set of responses to prevent the nude image from being shared.

However, age-based restrictions do not prevent adult accounts from being able to create, receive or share CSAM. Neither do these restrictions stop a child from using their device to share imagery through other distribution methods (e.g., encrypted environments). Device level protections are needed, alongside with platform level protections, to effectively tackle the threat of CSAM online.

By device level protections we mean:

- Nudity detection and blocking on children's devices
- Mandatory CSAM detection and blocking on all devices

3a. Device-level nudity detection and blocking on children's devices

Introducing device level protections would make a real difference to children and demonstrate Government's commitment to tackling online CSEA. Detection on children's devices addresses a specific set of harms which include 'self-generated' imagery[25] and grooming.

"I really thought this guy I was talking to was my age. We had been texting for a while, then it became sexting and then he asked for nudes. I sent some but he insisted I send more with my face in. He just kept asking and peer pressuring me so I just did it. Now

[23] Lucy Faithfull Foundation (2026). Project Intercept Impact Report [pdf]. p. 1.
[24] NCVPP (2025). 2024 National Analysis of Police-Recorded CSAE crimes report [pdf].
[25] The term 'self-generated' child sexual abuse as an inadequate and potentially misleading term which does not fully encompass the full range of factors often present within this imagery, and which appears to place the blame with the victim themselves. Children are not responsible for their own sexual abuse. Until a better term is found, however, we will continue to use the term 'self-generated' as, within the online safety and law enforcement sectors, it is well recognised.

he's told me he's 32! I don't want to talk to him anymore, but he has my pictures and is threatening to share them." - Girl, aged 16, Childline

In 2024, 91% of all IWF reports confirmed as containing CSAM included at least one self-generated image or video, and in 2025 this proportion decreased slightly to 85%.[26] Generating these images can be as a result of grooming, extortion, and coercion as outlined by IWF Analysts:

> *"Some children are heavily coerced into sexual behaviours by playful, flirtatious attention, games and emojis. Others are encouraged to exchange 'nude' imagery in what is disguised as a relationship or mutual sexual exchange between peers. We have also observed disturbing instances of humiliating sexual extortion, where a child is threatened with exposure if they do not comply with demands for sexual content.*
>
> *In so many cases, what starts as one sexual interaction between a child and another person can turn into hundreds of online child sexual images, posts, views and shares."*[27]

To mitigate the threat of children sharing nude images of themselves, or being exposed to this content, Government must urgently prioritise introducing device level protections that block the creation, sharing, and viewing of nude images on children's devices.

Creating child safe platforms must also be complimented with creating safer devices for children. We welcome the Government's pledge to work with technology companies to develop solutions to tackle image-based abuse and encourage companies to bring in on-device nudity detection filters.[28] Nudity detection technology needs to be introduced with appropriate safeguards in place for tackling issues around false positives and disclosures around safeguarding. Such tools already exist through in-house nudity detection technology at Meta, Apple, and Google (see section 1) and third-party technology such as SafeToNet's AI detection tool, Harm Block.[29]

3b. Considering mandatory CSAM detection and blocking on all devices

We encourage Government to introduce detection and blocking of child sexual abuse images and videos on all devices to prevent any user from being able to share this material. Whilst it is important that both known and novel CSAM are blocked at source, we encourage Government to prioritise introducing requirements for on-device detection of known CSAM as a start given its high accuracy in detection.

Detection on all devices prevents offender-facing harms, such as live streaming online CSEA, from falling under the radar. Livestreams can be abused by perpetrators to produce and share CSAM, as outlined in the IWF's 2018 one-off snapshot study.[30] CSEA livestreams can often

[26] IWF (2025). IWF Annual Data & Insights Report 2024 - Reports assessment [web].
[27] IWF (2026). Internet Watch Foundation Annual Report 2025 - 'Self-generated' imagery [web].
[28] Home Office and DSIT (2025). Protecting young people online at the heart of new VAWG strategy [news].
[29] SafeToNet (n.d.). Stop the camera. Block the Harm [web].
[30] IWF (2018). Trends in Online Child Sexual Exploitation: Examining the Distribution of Captures of Live-streamed Child Sexual Abuse [pdf].

happen on adult-to-adult accounts where the abuse of the child is facilitated by a third party.[31] Whilst adult livestreaming accounts can continue to facilitate this live and growing harm, Government's proposed age restrictions would not mitigate their threat.[32]

Livestreamed CSEA is a global threat that will require global interventions, as acknowledged in the Five Country Ministerial roundtable in September 2025.[33] The International Justice Mission (IJM) UK notes that in 2019 the National Crime Agency revealed the UK to be one of the top three global consumers of for-profit livestreamed child sexual abuse.[34]

According to IJM UK, livestreaming child abuse is one of the fastest growing forms of trafficking and often done at the hands of family members.[35] For Cassie,[36] a survivor of online child sexual exploitation, the abuse began at the age of 12. Cassie suffered this horrifying exploitation for almost five years until she was brought to safety.[37] Now, "Cassie is living with purpose. She is shaping a future where children are safe, survivors are empowered, and all are free."[38] Survivors like Cassie are urgently calling for governments to act now and require companies to proactively detect and disrupt livestreamed CSEA.[39]

4. Adopting a risk-based approach to restrictions on features and functionalities

Any restrictions on children's online activity should be grounded in evidence of risk, not arbitrary platform categorisation. High-risk features and functionalities, including stranger messaging, live streaming with unknown adults, and nudity sharing, should be age-restricted in a graduated way (see section 6b for our views on a principle-based age-appropriate experiences). A model similar to the British Board of Film Classification (BBFC) film classification system should be implemented that includes an independently set criteria, applied consistently across all services, with enforcement against non-compliance. The BBFC's classification system provides age ratings and content advice on films and other audio visual content to help children and families determine whether the content being viewed is right for their family or whether it needs to be avoided. Every 4-5 years, the BBFC consults thousands of people across the country in order to ensure that its Classification Guidelines remain in step with societal expectations.[40] This form of classification avoids the "whack-a-mole" dynamic of platform-by-platform or age-by-age bans. New standards should be introduced which provide clear guidance on how services must deliver age-appropriate experiences for children online.

5. OSA gaps and Ofcom implementation

---

[31] Cubitt, T., Napier, S., and Brown, R., (2025). Financial risk indicators of child sexual abuse live streaming: A proof of concept prediction model [pdf]. *Trends & issues in crime and criminal justice.*
[32] International Justice Mission (2023). Scale of harm: Research method, findings and recommendations [pdf].
[33] Home Office (2025). Five Country Ministerial 2025: Communiqué [web].
[34] The National Archives (2022). The Report of the Independent Inquiry into Child Sexual Abuse [pdf].
[35] IJM UK (n.d.). What is livestreamed child abuse? [web].
[36] Pseudonym.
[37] IJM Australia (2025). Cassie's* Story [video].
[38] IJM (2025). Cassie: From Survivor to Child Protection Advocate [web].
[39] IJM Australia (2025). Cassie's* Story [video].
[40] BBFC (2024). Classification Guidelines [pdf].

This section has been endorsed by the Online Safety Act Network.

6. Regulating AI chatbots and companion services

The risk and severity of AI generated child sexual abuse imagery continues to grow, as reported by the Internet Watch Foundation in their 2026 report *Harm without limits: AI child sexual abuse material through the eyes of our Analysts*.[41] The IWF reports that AI chatbot services have been found to share AI-generated CSAM and encourage users to act out simulated CSEA scenarios.[42] Currently, chatbot services are not adequately covered by the Online Safety Act, which only includes user-to-user, search and pornographic services. This means the legal liability for chatbot companies remains unclear with coverage partial and uneven. Government must ensure all AI chatbots are adequately regulated under the Online Safety Act and take action to tackle chatbots which assist or encourage child sexual abuse or simulate the offence of sexual communication with a child.

Chatbots should also be required to signpost users to support services where conversations involve references to child sexual abuse. Specifically, where a chatbot is asked for information by an end-user about child sexual abuse imagery, the chatbot should be required to refer the end-user to Lucy Faithful Foundation's Stop It Now and/or the IWF and NSPCC's Report Remove tool as appropriate. Where a chatbot is asked about the experience of child abuse, the chatbot should be required to refer the end-user to urgent support.

The threat of CSEA extends across gaming platforms, messaging services, AI tools, and encrypted environments. Any legislative or regulatory intervention must reflect this wider landscape and be designed to address the full range of CSEA threats, not only those occurring on mainstream social media. With this in mind, we welcome the introduction of standalone OSA duties through Section 216A of the Crime and Policing Act (2026). The Government must use its powers to amend the Online Safety Act to include AI chatbots in its scope, as enabled by the Crime and Policing Act, with Ofcom then utilising its existing powers to require enforceability of minimum ages within the Children's Code. To ensure AI chatbot services are fully in scope of regulatory duties, there must be further legislative amendments needed which consider: risk assessments and transparency notices for AI chatbots; legal requirements for robust product testing; penalties for failing to mitigate and manage risks; and sign-posting to provide victim support or behavioural changes.

7. Requiring effective age assurance across services

The majority of technical interventions require reliable identification of child users. To achieve this, we are specifically calling for:

- Establishing enforceable minimum age requirements
- Introducing a legislative framework for principle-based age-appropriate experiences

[41] IWF (2026). Harm without limits: AI child sexual abuse material through the eyes of our Analysts [pdf].
[42] Ibid.

6a. Establishing enforceable minimum age requirements

Establishing enforceable minimum age requirements is an essential part of the age assurance process. Platforms should not be permitted to self-determine the minimum age-appropriate for their services as this approach leads to ineffective implementation. Ofcom's 2025 *Children and Parents: Media Use and Attitudes Report* found that children as young as 3 to 5 are using online services.[43] It is clear that enforcing minimum age requirements will be an essential tool to achieve the ambitions of the consultation.

> *"I was using Wizz app with my friends when someone (who I thought was 13) cut in and pm'd [private messaged] me. They asked me for a nude photo and threatened me, saying that they knew where I lived, so I sent it to them. I've spoken to my parents about it and I've blocked them. I've also reported it to CEOP after I spoke to Childline about it. It's making me feel anxious and sick." - Girl, aged 11, Childline*

Minimum age thresholds should be independently established and enforceable using a model similar to the BBFC classification system (see section 4 'Adopting a risk-based approach to restrictions on features and functionalities'). Ofcom has acknowledged that minimum age enforcement is not explicitly mandated across the Online Safety Act.[44] The ICO's Children's Code already expects services likely to be accessed by children to take a risk-based approach to age assurance.[45] In the joint statement by Ofcom and the ICO on age assurance, it is also re-stressed that platforms with minimum age requirements should move beyond self-declaration and use effective technological solutions for enforcement.[46] The joint statement also reiterates that services must enforce their minimum age rules using highly effective age assurance.

6b. Introducing a legislative framework for principle-based age-appropriate experiences

Simply restricting access to sites for certain age groups does not go far enough to ensure children have age-appropriate experiences. By age-appropriate experiences we mean where age and stage of development of the user is meaningfully considered as part of platform design. Ofcom must provide a regulatory framework for principle-based age-appropriate experiences which recognises that children of different ages should not have identical online experiences. This should be in conjunction with Government, who should introduce a clear statutory framework for Ofcom to then implement said framework. Effective age assurance, properly implemented, creates the infrastructure for graduated, risk-based protections and opens new opportunities for enforcement (to learn more about risk-based protections see section 4). Delivering a principle-based age-appropriate experience will be central to child safeguarding, and not an optional extra. The Children's Coalition for Online Safety further elaborates on how Government can adopt a principles-based approach to age-appropriate online experience.[47]

---

[43] Ofcom (2025). Children and Parents: Media Use and Attitudes Report [pdf].
[44] Ofcom (2026). Keep underage children off your platforms, Ofcom tells tech firms [web].
[45] ICO (2022). Age appropriate design: a code of practice for online services [web].
[46] DRCF (2026). Age Assurance: A Joint Statement by Ofcom and the Information Commissioner's Office [pdf].
[47] Children's Coalition for Online Safety (2026). Joint Statement on the Children's Online Safety Regime [pdf].

8. Strengthening the Online Safety Act

An outcomes-based overarching duty of care must be introduced which places responsibility on platforms for the risks their services create. The OSA Network's 10-point plan[48] outlines the urgent amendments needed to strengthen the Act, some of which are outlined below. These interventions aim to go with the grain of legislation and shift Ofcom's implementation of the regime towards a more risk-based, outcome-focused approach that puts the onus on services, rather than the regulator, to keep users safe.

7a. Stopping unsafe products from reaching the market

Regulated services must be required to take reasonable steps to reduce the risk of harm to users as identified in their own OSA risk assessments. The Online Safety Act's "clear and detailed" and "technically feasible" conditions for measures in Ofcom's codes of practice should be revised to allow for more stretching, outcome-focused measures. In addition, the 'safe harbour' provision in the Online Safety Act, which currently deems companies to be meeting their safety duties as long as they are in compliance with the Codes, must be removed. Safe harbour disincentivises companies to go beyond meeting the minimum requirements listed in the Codes and fails to raise the bar for ambitious safeguarding measures and innovation across the sector.

7b. Ensuring all aspects of products are safe by design

Ofcom's approach to implementation has focused primarily on downstream measures, which deal with the harm after it has happened, rather than on the upstream requirement – as set out in the first section of the OSA - that services need to be "safe by design." This is discussed further in the OSA Network's 10-point action plan.[49] A statutory definition of safety by design must be inserted into the Act to make clear to Ofcom and regulated services what Parliament intended. The current lack of definition, alongside codes that rely largely on evidence-based and prescriptive measures, has led to a series of codes of practice that are narrow and limited to ex-post actions. Secondary legislation should be laid to require Ofcom to produce a cross-cutting safety by design code of practice which would sit beneath the existing content-based codes and operationalise the intention of the Act. The Online Safety Act Network has worked with key partners, including signatories of this document, to produce a draft Code of Practice to demonstrate what good "safety by design" looks like.[50]

7c. Strengthening user protection

We are calling for three specific amendments needed to the Online Safety Act to ensure that the baseline of user protections intended by Parliament is fully delivered:

[48] OSA Network (2025). Strengthening the Online Safety Act: A ten-point plan for Government [pdf].
[49] Ibid. p. 5
[50] OSA Network (2026). Safety by Design Code of Practice [pdf].

- **Introducing a statutory stay-down provision.** The Online Safety Act should make clear that once illegal content or imagery is removed, any further shares of the same material must also be automatically removed, without requiring a repeated moderation process.
- **Amending Schedule 11 of the Online Safety Act to ensure all high-risk platforms, regardless of their size, are in scope of existing illegal content duties.** Parliament amended the Online Safety Act to allow services to be designated as Category 1 on the basis of size or risk and thus subject to additional duties, ensuring that small but high-risk platforms fall within scope of the enhanced duties. However, Ofcom's interpretation of the Online Safety Act, and the Government's subsequent regulations, have not taken this into account. Government should re-visit the categorisation to ensure all high risk-platforms, regardless of size, are in scope of enhanced duties.
- **Setting minimum standards for the Terms of Service for Category 1 services.** These minimum standards should mirror the criteria in the user-empowerment duties under section 15 of the Online Safety Act. The minimum standards must also include a 'no rolling back' requirement. At present, services can dilute protections for users with no sanctions under the OSA.

7d. Enabling wider protections against CSEA harms

The Online Safety Act contains gaps in tackling CSEA and related harms that require substantive Government action, which is independent to Ofcom's implementation of the Act. The interventions below would address these gaps.

- **Including an overarching duty of care in the Online Safety Act.** An overarching duty of care should be inserted into the legislative framework to require a more systemic approach to risk assessment and mitigation from regulated services, including in relation to emerging harms arising from new products and services.
- **Placing a duty on Ofcom to cooperate with domestic regulators and placing the DRCF on statutory footing.** A coordinated regulatory approach is essential to ensure that emerging harms arising from new products and services do not fall between institutional responsibilities, and that the protections Parliament intended when passing the Online Safety Act are delivered in practice across the full range of regulators with relevant powers. A duty should be imposed on Ofcom to cooperate with other domestic regulators, by mirroring the existing duty within the Online Safety Act regarding overseas regulators. This gap can begin to be rectified by placing the Digital Regulation Cooperation Forum (DRCF) on a statutory footing.

Conclusion

At the heart of a safer digital environment for children online is an environment that is free from the risks of child sexual abuse and exploitation. The measures we have proposed can radically disrupt the cycle of CSEA, and their implementation within the online safety regime is easily achievable. No single intervention alone can tackle the existing, evolving, or emerging CSEA

harms. For the past few months, age restrictions have been the central focus of Government activity around online safety. However, it is important to note that with the introduction of age-based restrictions the fight to keep children safe online is not yet won. We urge the swift introduction of our proposed suite of interventions to deliver monumental change for children safety online.

Appendix

Growing up in the online world: Collective recommendations for tackling the risks of online child sexual abuse and exploitation - Appendix

The table indicates the signatories' views on the likely contribution each intervention has in reducing the specific harms outlined below. As mentioned in the response above, the table does not suggest that any single intervention is sufficient on its own and the effects will vary depending on a number of technical factors including, but not limited to, service design, age assurance, cross-platform displacement, enforcement and user behaviour.

The relationship between the harm and the intervention is noted based on two categories: level of contribution for tackling the harm or level of enablement in facilitating the harm.

The scale for both categories is as follows:

- Level of contribution for tackling the harm – direct, partial, limited, none
- Level of enablement for facilitating the harm – strong, medium, low

Intervention	Primary mechanism	Self-generated indecent imagery	Grooming	Sexual extortion	Known CSAM recirculation and/or sharing	CSAM livestreaming	Key limitations/caveats
Social media ban	Restricts access to certain mainstream platforms	Partial	Partial	Partial	Limited	Limited	• May reduce exposure on some platforms, but risks displacement to other services like messaging, gaming, encrypted and potentially adult spaces • Does not directly address adult offender networks • May reduce children reporting harms if they are not supposed to be in those spaces
Device level nudity detection and blocking on children's devices	Prevents creation, sharing and viewing of nude imagery on child accounts and devices	Direct	Limited	Partial	Limited	Limited	• Strongest for child facing image harms • Depends on: (1) accurate age identification, (2) proportionality safeguards and (3)

							routes for review and appeal
Risk-based restrictions on high-risk features and functionalities by age	Limits risky features including stranger messaging and contactability; forwarding; and (potentially) livestreaming	Direct	Direct	Partial - Direct	Limited	Partial	• Depends on robust age assurance and consistent application across services (not just social media platforms in scope of regulation) • Likely strongest when targeted at specific high-risk features
On-device blocking of known CSAM	Prevents access to or storage of known abuse material	Limited	Limited	Limited	Direct	Partial	• Stronger for known CSAM than for novel material • Does not by itself address other forms of OCSEA such as solicitation and grooming or wider non-image harms
Upload prevention/pre-encryption	Prevents known CSAM being uploaded or shared which includes	Limited	Limited	Limited	Direct	Partial	• Important for known CSAM circulation • Does not by itself address grooming, abuse that is text-

detection of known CSAM	encrypted environments						based, or newly created/novel imagery
Effective minimum age requirements and age assurance	Creates infrastructure for differentiated protections and access controls	Enabling	Enabling	Enabling	Limited	Limited	• Better understood as an enabling condition than a standalone solution • Effectiveness depends on moving beyond self-declaration and age-estimation
Deterrence messaging and offender signposting	Introduces friction and warnings as well as potential pathways to support for users displaying concerning behaviour	Limited	Limited	Partial	Partial	Limited	• Best seen as a complementary prevention and disruption measure rather than a primary safeguarding control
Regulation of AI chatbots and companion services	Prevents harmful simulation, generation, encouragement or facilitation of abuse	Limited	Partial	Partial	Limited	Limited	• Particularly relevant to emerging harms • Needs clearer duties around harmful outputs • Needs more guidance on behavioural

							encouragement and support signposting