RESPONSE TO OFCOM'S ADDITIONAL SAFETY MEASURES CONSULTATION: INDIVIDUAL MEASURES

OCTOBER 2025

Introduction

This paper brings together our analysis on the individual proposals in Ofcom's Additional Safety Measures consultation, responding to the specific questions posed by Ofcom, and will be submitted as part of our response. We are grateful for the contributions and inputs of members of the Network to our analysis. This response should be read in conjunction with our covering paper which draws out some of the more thematic issues that arise from the proposals and our analysis on the "technical feasibility" proviso.

Livestreaming

Do you have further evidence regarding the harms and risks to users from livestreamed illegal content or content harmful to children, or harms and risks to children from broadcasting livestreams?

1. As highlighted by the OSA Network and numerous other civil society organisations in their responses to both the Illegal Harms Code of Practice and the Protection of Children's Code of Practice, livestreaming - facilitating the connection of users to one another in real time - can pose severe risks to user safety, particularly children. Ofcom's own evidence, compiled in the risk register for each consultation, underlined this yet measures to address the clear risk are currently missing from Ofcom's codes of practice. It is a feature that can lead to multiple harms, with particular risks for children and vulnerable users and one where the principle of "safety by design" (as we set out in our covering paper to this response) is acutely relevant.

2. There is already significant evidence of harm caused by livestreamed content, including well-documented cases such as the Buffalo terrorist attack[1] and livestreamed suicide attempts[2].

[1] https://www.ofcom.org.uk/siteassets/resources/documents/research-and-data/online-research/other/buffalo-attack/the-buffalo-attack-implications-for-online-safety.pdf?v=328573

[2] https://www.theguardian.com/news/2017/may/21/facebook-users-livestream-self-harm-leaked-documents

Research carried out by the NSPCC found that 6% of children who livestream have been asked to remove their clothing[3], and Internet Matters found that almost a third of users aged 11 to 13 are using livestream, despite the average minimum age for joining most large platforms being 13[4].

3. To recognise this harm, Ofcom is proposing two additional safety measures aimed at improving all users safety whilst using live-streaming technology. While welcome, these are, however, *ex-post* measures - improving user reporting where a livestream "depicts imminent physical harm" and requiring better content moderation and action in real time - rather than upstream measures which will make livestreaming "safer by design" as Ofcom claims on p27. In addition, Ofcom is suggesting additional measures for platforms to protect child users of livestreaming technology, which is also a welcome step forward but which - as we set out below - could go further.

4. Ofcom has not been clear about what further evidence it requires to justify stronger action, despite already including livestreaming in the Register of Risks published in December 2024. Here they acknowledge that:

   There are many examples of terrorists livestreaming attacks, this can in turn incite further violence. The use of livestreaming remains a persistent feature of farright lone attackers, many of whom directly reference and copy aspects of previous attacks. Similarly, perpetrators can exploit livestreaming functionality when abusing children online.

   We know that many civil society organisations will submit further evidence to this consultation but the requirement to do so, given what Ofcom already knows about the significant risks associated with livestreaming, is frustrating given the significant gap that remains between Ofcom's assessment of risk and its recommendation of measures to mitigate that risk. (See for example our updated table of measures.)

5. We also note that the regulator is "working to obtain further insights *on the benefits* and risks" of livestreaming "from children and those who care for them, experts and those with lived experience of online harms". Ofcom's protection of children risk register, published in May 2024, contained the following assessment: "For example, while livestreaming *can be a risk factor for several kinds of harm to children (as it can allow the real-time sharing of harmful content such as suicide and self-harm content*), it also allows for real-time updates in news, and can provide children with up-to-date tutorial videos and advice or encourage creativity in streaming content. These considerations are a key part of the analysis underpinning our Protection of Children Codes measures." (p21) The evidence of the risks outweighed the benefits 18 months ago and, for the regulator to still be asking for more evidence on benefits, while refusing to act as

[3] https://learning.nspcc.org.uk/media/1559/livestreaming-video-chatting-nspcc-snapshot-2.pdf

[4] https://www.internetmatters.org/hub/research/rise-of-the-streamager-nearly-a-third-of-11-to-13-year-old-are-broadcasting-themselves-live-over-the-internet/

comprehensively as it could on the well-evidenced risks, seems perverse.

6. Ofcom must do much more to demonstrate that it is acting in accordance with the overarching duties in the Online Safety Act (section 1) that regulated services must be "safe by design", especially given that higher standards must be in place for children. Our previous concerns about Ofcom codifying a lowest common denominator approach remain here: Ofcom mentions almost in passing at the end of the chapter on livestreaming that some services have already taken steps to prevent children from livestreaming entirely "as a matter of their own service design", yet they have not collected evidence from these services as to how they implemented the measure and the impact it has had. This would have been useful to help the regulator put the case in this consultation for an option for a default ban for children (or a partial ban depending on age) and set the bar higher for regulated services who have not taken such steps. Instead, the regulator asks for yet more "compelling evidence" from third parties - building in yet more delays with further consultation requirements - rather than providing a suite of options, including existing industry practice, in this one.

7. Whilst the risk register already identifies livestreaming as a high-risk area for multiple types of harm, this is not being translated into preventative requirements such as live-feed delays or mandatory proactive moderation, indicating that Ofcom is prioritising "evidence of what works" rather than "evidence of harm"; and is operating on the basis that absence of evidence that something works means that it does not work. Ofcom must be transparent about how it evaluates evidence and should publish a clear framework showing how evidence of harm informs regulatory decisions. In relation to other measures (proactive tech), Ofcom has adopted a principles-based approach allowing providers to identify what works (see para 8.12). Surely a principles based approach could additionally be adopted here, to help fill the gap in evidence that Ofcom seems to think that it has? We would welcome clarification from Ofcom about this inconsistency in approach between these measures.

Livestreaming: proposals for all users

Do you agree with our proposals? Please provide your reasoning, and if possible, provide supporting evidence.

8. We welcome the fact that Ofcom is introducing measures to address the well-evidenced risks from livestreaming but we have concerns that these do not go far enough. Indeed, Ofcom presents these as a "minimum set of measures" to "set a baseline to protect users". This is not ambitious enough. We set out our analysis and some suggestions for improvements below.

Application to all users

9. D17 - the requirement “for a mechanism to enable users to report that a livestream contains content that depicts the risk of imminent physical harm” - claims to apply to “all users,” yet is limited to U2U services at medium to high risk of terrorism, CSAM, suicide, and hate, stalking and abuse offences. The concerns that arise in relation to these offences vary. Some relate to the impact on the person making the content (e.g. suicide content), and some relate to the impact of the content on the viewers (eg terrorism content). Other serious offences, such as animal cruelty, trafficking, or extreme violence, are not included, despite being high-risk and capable of causing imminent public harm if livestreamed. Indeed, in paragraph 4.30 Ofcom have listed animal cruelty as one of the listed harms associated with livestream: it is excluded from recommendation D17 but included in C16 (the requirement for human moderators to be available whenever users can livestream). The exclusion of these categories creates inconsistencies in protection of users and leaves victims of other serious crimes without adequate safeguards.

Effective response measures

10. There is significant ambiguity in how Ofcom defines “imminent physical harm.” It is unclear whether the focus is on preventing harm to an individual currently in danger during a livestreamed event, or on preventing the spread of harmful content after the fact. This lack of clarity creates uncertainty for services and may delay urgent interventions. There are a number of other circumstances where livestreaming can be a signal for risk of imminent harm and, given that the crisis response proposals on which Ofcom is consulting are relatively light-touch, we have concerns that Ofcom is leaving a gap in protections. For example, livestreaming of violent protest was a factor in the escalation of the riots that followed the events in Southport (discussed by Ofcom in relation to crisis management, para 20.7), but this is not adequately covered by either measure (see below for our response). There are also boundary issues relating to the potential imminent harm that may be being broadcast on a livestream: for example, someone loading a gun outside a place of worship might be terrorism or it might be violence. If a user is faced with a number of reporting categories that do not fit with their evaluation of the imminent physical harm, then there is more risk of non-reporting than over-reporting.

11. Paragraph 5.15 of the consultation only requires that users be able to report content showing imminent physical harm, such as a live terrorist attack or suicide attempt. However, it does not require services to act rapidly or to integrate these reports into a wider moderation response system. There is concern that Ofcom has chosen narrow measures. There is a reference (at 5.17) to services that already have a category to allow users to report a livestream showing imminent physical harm but Ofcom has made no attempt to evaluate the effectiveness of these existing measures nor to require those services to go further in responding to risks, while bringing others up to that minimum baseline.

12. Even when users report imminent harm, there is no guarantee that services will respond in time to prevent further harm. Nor does Ofcom (at 5.19) provide guidance or obligations on reporting

this risk of harm to emergency services or other public bodies. While some large services will presumably have well-established mechanisms in place, it would seem reasonable to assume that services that do not already have a measure in place to deal with reports relating to imminently harmful livestreams will not.

13. Industry feedback has highlighted that providing reporting buttons on their own are meaningless unless backed by rapid human moderation and escalation pathways. Ofcom's current proposal leaves these processes open-ended, meaning services could meet the letter of the regulation without dedicating sufficient resources to real-time intervention. Given that these proposals will require resources to be allocated by tech platforms, Ofcom should make it clear that these resources are additional to those required for standard content moderation under their OSA duties.

14. Additionally the concept of a "large audience" is poorly defined. It's possible to broadcast to a large audience with very little engagement, while a smaller but more active audience on a platform that falls below the threshold set by Ofcom could cause greater harm, particularly (for example) if it is dedicated to a potentially harmful activity or topic, like a suicide or self-harm site.

Response vs. prevention

15. Ofcom's proposals focus on responding to harm after it occurs and content moderation rather than preventing it in the first place. There is no requirement for live-feed delays, which are standard practice in traditional broadcasting, to prevent harmful or illegal content from being aired in real time. Safety-by-design means including proactive measures such as time-delay buffers and real-time risk assessment. There is plenty of guidance available to broadcasters on this topic[5][6].

Do you consider that there are alternative measures which would materially reduce the risks to users from livestreaming such as preventive safety by design frictions, prompts or restrictions? If so, please detail them and provide evidence on the costs and efficacy.

16. Ofcom notes (at 4.16) that one of the risks for livestream broadcasters is the potential for harm when acting "in the moment", and the impact that financial reward, validation or recognition can have on those decisions. Yet elsewhere, in its discussion of negative impacts for users if the proposed measures are brought in, it assesses that 16 or 17 year olds may "experience reduced financial gain from livestreaming through our proposed limitations on in-service gifting" or that income from other sources such as subscriptions and sponsorships may decline due to "reduced visible engagement with their content". While Ofcom is not arguing in this section that their

[5] https://www.bbc.co.uk/editorialguidelines/guidelines/harm-and-offence/guidelines#liveoutput
[6] https://www.channel4.com/4compliance/compliance-guidelines/live-programme-guidelines

existing proposals should be rejected but merely setting out one of the impacts of restrictions on livesteaming, the lack of connection between the two statements is troubling: the prospect of financial gain increases the risk of harm (especially as the rules to protect children in the working environment[7] and safeguard their education do not apply online) yet the reduction of this risk is seen as a "negative impact" on teenagers. Furthermore, in its discussion on the indirect effects on service providers (6.47), Ofcom highlights that the "number of child broadcasters on livestreaming services may reduce significantly as livestreaming becomes less popular once interaction functionalities are removed" and that "the number of viewers on children's livestreams may also reduce due to the lack of ability to interact with those undertaking the livestream", both of which could "reduce services' revenue". On the basis of this analysis, there would clearly be a significant financial impact for services if Ofcom went further and recommended that livestreaming was turned off by default for children (whether under-16 or under-18) and it is difficult not to take the view that Ofcom's reluctance to recommend this - asking for more "compelling evidence" as a basis for action - is due to the perceived backlash from services who care more about profit than child safety.

17. More broadly, we would recommend that Ofcom consider a greater array of ex-post features - e.g. borrowing from broadcasting good practice and building more delay into a live stream as a feature. While the limits this would place on "real-time" user interaction could potentially have more impact on communities on some platforms (eg those that host livestreams for gamers) than others, it should be part of the risk-assessment consideration for platforms: balancing the ability to intervene and stop problematic live streams before harm occurs and disrupt the impact of "in the moment" pack behaviour rather than waiting for the harm to be reported before action is (belatedly) taken.

Livestreaming: proposals to protect children

Do you agree with our proposals?

18. Section 6 of Ofcom's consultation focuses on livestreaming by children and introduces new safety proposals under measure ICUF3. There is much to welcome in the proposed measures, which recognise the increased vulnerability faced by children using livestreaming services, and the need for strong reporting mechanisms to allow users to flag imminent harm. Action on livestreaming has long been called for by child protection charities due to the significant risks of harm that emerge for children.

19. It is also important to note that safeguards in the draft proposals apply only to 'one-to-many' livestreams and not to 'many-to-many', creating a loophole. Most live-streaming platforms

[7] https://www.cbc.ca/news/canada/london/study-warns-about-dangers-of-kidfluencers-kids-and-content-creators-have-a-different-opinion-1.7628919

enable users to add a cohost(s) to their stream, meaning one-to-many livestreams can easily be turned into many-to-many livestreams and bypass the need for ICUF3 protections without reducing the risk children face. A more appropriate way to define where ICUF3 protections should apply would be to consider service risk levels. On platforms assessed as high risk for grooming and image-based CSAM, the ICUF3 protections should be in place to offer additional protections to child users, regardless of whether it is used for on-to-many or many-to-many streaming.

Age appropriate

20. Livestreaming is inherently high risk, providing a direct, unmoderated line of communication between children and between children and adults who may wish to cause them harm. While an outright ban might be one approach to mitigating this risk - and Ofcom mention in their consideration of potential next steps that preventing children from livestreaming entirely is a "step that some services have already taken as a matter of their own service design" - civil society organisations have highlighted that there is a difference between the vulnerabilities of children under 16 and those aged 16 to 17. One approach would be to have access to livestreaming functionality off by default for children under 16, as outlined by NSPCC in a recent evidence session organised by the Communications and Digital Committee[8]. For older teenagers aged 16 to 17 there is an argument for a more nuanced approach to 'age appropriate' restrictions, recognising their greater autonomy than younger children. This might include giving them the option to enable livestreaming in tightly controlled circumstances once safety measures have been proven effective.

Safety-by-Design

21. As we set out in our response to the measures proposed for all users, we believe that there should be more of an onus on platforms to build in protective systems or frictions that disrupt harmful behaviours before they occur. A strong understanding of safety-by-design would mean that where livestreaming cannot be delivered safely it shouldn't be in place. See more on this in our detailed paper on Safety by Design[9].

22. There is also a danger that the proposed measures are a reflection of industry norms rather than pushing innovation. Indeed, as we note above, given that some services do not allow children to use the livestreaming functionality, such an approach would not even be that innovative.

[8] https://committees.parliament.uk/oralevidence/16484/pdf/
[9] https://www.onlinesafetyact.net/analysis/safety-by-design/

Process over outcomes

23. We note that Ofcom's measures focus on process**.** While this is an essential element to developing safer services, outcome-orientated objectives are also required especially given the higher level of protection required for children by the Act. It is vital that the regulation demands clarity from tech platforms about what actions platforms they have taken when harm is identified. From a law enforcement perspective, there is a need for real-time escalation pathways to law enforcement when imminent harm is identified by platforms.

Proportionality

24. Platforms have a clear duty to address risks such as CSAM, whatever their size. Allowing a lower standard in one area or for some types of services simply drives bad actors to smaller, less regulated services. The onus should be on platforms to innovate safety measures, not to claim exemptions because they lack technical solutions. For a discussion of technical feasibility and its relation to proportionality see here. We would also draw Ofcom's attention to this analysis in relation to its own discussion of technical feasibility in paras 6.28 to 6.31 and to the decision to include the proviso in the measure on screen capturing and screen recording. We do not see why it is not appropriate to recommend a measure for no content capture for all video streams: while it is right that those who are determined to subvert it will find ways round it, introducing this friction might reduce the risk of "in the moment" reactions to eg terrorism content spreading

Proactive technology

Do you agree with our proposals? Please provide your reasoning, and if possible, provide supporting evidence

25. We broadly support the move towards requiring proactive technology as a safety-by-design approach to user safety. The principles-based approach taken by Ofcom allows platforms to adapt over time as technology evolves, and it is welcome to see the onus being placed on service providers to exercise a duty of care to the children who are using their platforms. We also welcome the inclusion of review of existing proactive technology used by services; this mitigates the risk that legacy tech of insufficiently high standard bakes in inaccuracy or bias and supports an approach of ongoing review of such tech. However we note several areas in which there could be more ambition in Ofcom's approach, as well as the need for careful guardrails around tech interventions that are still in their nascency. We would also flag here (and below) the risks around separating out measures on hashmatching, which is a distinct type of proactive technology, for CSAM, terrorism and intimate image abuse (IIA) but then excluding IIA, but not the other content, from the broader proactive technology measure.

26. We also support points raised in EVAW's response:

In relation to proactive technology and hashmatching the focus is almost exclusively on detection. While detection can assist in prevention, it cannot and should not be a stand in for it, as in these instances the harm itself has largely already occurred. (Though we recognise the reduction in proliferation). Ofcom could go further at developing requirements on platforms that aren't reliant solely on detection.

Technically feasible

27. We note the Ofcom clarifies that it does "not consider that it would be technically infeasible to implement proactive technology merely because to do so would require some changes to be made to the design and/or operation of the service" (para 9.55) Nonetheless, the technically feasible test risks allowing providers to evade compliance by claiming cost or complexity especially because a consequence of including proviso wording in the measure is that this has the effect of allowing the service providers – at least in the first instance – to determine whether they need to comply with the measure or not. While Ofcom will ultimately be able to determine the appropriateness of that decision in a given case, a key concern is the approach of Ofcom to monitoring self-declared technical infeasibility, recognised in para 9.58. What would be the trigger for the investigation and when? While we welcome Ofcom's recognition that technical limitations are not a "once and for all" determination (para 9.58), and agree that technical feasibility must be kept under review, again there are questions about the nature and frequency of oversight of that review process. We set out these concerns in more detail, and particularly the relationship between technical infeasibility and cost, and make a series of recommendations in this commentary piece While we also note that Ofcom points out that a change to make a technology technically infeasible would trigger a risk assessment (para 9.59), given the current content of the Codes and the safe harbour provisions, it is unclear whether services would have to take mitigating steps. As set out in our paper, we suggest that some form of no-rollback requirement in relation to user safety standards should be introduced to counter this possibility.

Do you agree with the harms currently in scope of these measures? Are there any additional harms that these measures should capture? Please provide the underlying arguments and evidence that support your views, including evidence regarding the availability of accurate and effective proactive technology.

28. Whilst we broadly agree with the harms Ofcom proposes to bring into scope for proactive technology, it is unclear why intimate image abuse has not been included in this general section on proactive technologies. It is a priority illegal offence with profound consequences for victims. Moreover, it is mentioned in relation to a specific type of proactive technology: hashmatching. While relevant technology might not yet be available, it is possible that new tech would become available (eg extending network analysis techniques to determine those engaged in "collection" of non-consensual intimate images - and Ofcom recognises collector culture in para 11.6). The measures should not be frozen in time (especially in the light of Ofcom's points in para 9.77-78). We also note the potential risk of only focusing this measure on new content, and not

pre-existing material (with the exception of CSAM) (para 9.73), which risks leaving older illegal content circulating: we note that this is particularly a concern in relation to fraud. There does not appear to be any justification for this (given that privacy concerns should have been addressed at time of the decision to adopt and deploy the technology). Ofcom must commit to an iterative, evidence-driven expansion process that keeps pace with emerging forms of harm.

Do you agree with who we propose should implement these measures? Are there any other services that should be captured for some or all of the relevant harms?

29. The proposed risk-based targeting of large multi-risk platforms, high-risk services with more than 700,000 UK monthly users, file-storage services at high CSAM risk, and all services identifying a high grooming risk is an appropriate starting point, however there is still room for small but risky sites to fall through the gaps. The size threshold and tests should be refined so that platforms with disproportionately large child user bases or inherently risky features, such as livestreaming, direct messaging and gifts, cannot avoid obligations by citing overall user numbers. Ofcom should therefore include criteria that capture services with high child proportions or particular functionalities.

30. Given the inclusion of the technically feasible clause, Ofcom must allocate resources to test provider claims of infeasibility against a transparent technical framework to avoid inconsistent outcomes where different services reach divergent judgments on the same technology.

Amendments to illegal content judgements guidance for child sexual abuse material

Do you agree with our proposal? Please provide your reasoning, and if possible, provide supporting evidence.

31. We welcome the changes to the ICJG. This change is important given the limitations on detection as a result of the use of some technologies in a service's architecture. However we want to stress the importance of design in being able to assess illegality. For example, some platforms use metadata and can view some reported content when flagged. We therefore think that platforms should be required to share reported data to moderation teams.

Do you agree with our assessment of the impacts (including costs) associated with this proposal? please provide any relevant evidence which supports your position

32. We note the rights assessment accounts for the possibility of knock-on interferences with speech because of an incorrect assessment about a first piece of content (para 10.15) but that (amongst other considerations) it addresses a very serious harm. We do not dispute that the harm in issue is serious, but it is also a rights violation - of Article 8 and arguably Article 3 rights of the victim.

The fact that there is a rights conflict in issue means that a "fair balance" should be sought rather than adopting the three-stage test used when there is an interference with rights in support of more general public interests. Such an approach only reaffirms Ofcom's conclusion (para 10.19).

Perceptual hash matching for intimate image abuse

Do you agree with our proposals? Please provide your reasoning, and if possible, provide supporting evidence.

33. The proposals set out by Ofcom are a positive step towards ensuring survivors of intimate-image abuse have access to swift support from service providers when they have experienced abuse on their platforms. Civil society organisations such as the Revenge Porn Helpline, the UK Safer Internet Centre and EVAW have campaigned for these measures to be included in the Act, and we welcome the action taken by Ofcom.

34. We defer to their assessment of the effectiveness of the proposal but would also suggest that there are some areas that require clarification from Ofcom. As such, we echo and endorse the UK Safer Internet Centre's response:

    Industry hash sharing, combined with NGO insertion and survivor-led hashing, creates a multi-layered, future-proof system that balances privacy, accuracy, and speed. It enables platforms to act decisively, even in high-harm scenarios where survivors cannot participate directly. It also supports smaller platforms with limited moderation capacity, allowing them to prioritise verified hashes and respond confidently to known threats.

    Given the scale and severity of NCII, and the clear technical feasibility of hash sharing, it is reasonable to expect that platforms, particularly those already in scope under the Online Safety Act should implement this capability as part of their broader duty to mitigate illegal content. The infrastructure exists, the privacy safeguards are robust, and the public interest in preventing re-victimisation is compelling. Endorsing industry hash sharing as a regulatory expectation would ensure consistency across services and deliver meaningful protection for victims at scale.

    This evolution reflects StopNCII.org's commitment to continuous improvement and partnership. It strengthens the ecosystem's ability to prevent NCII, supports victims more comprehensively, and aligns directly with Ofcom's objectives under the Online Safety Act. We recommend that Ofcom formally endorse industry hash sharing as best practice and consider its inclusion as a regulatory obligation for in-scope services.

35. We also support EVAW's response:

Survivors should have agency over whether their images are added to a third-party hash database. Ofcom should require: (a) informed consent processes or consent-substitute protections where consent cannot be obtained, (b) trauma-informed victim support contact points, and (c) rapid removal and appeals processes that prioritise survivor safety and privacy.

We also believe that deepfakes are not sufficiently dealt with under the proposals. They should be more explicitly included in both the Hashmatching requirements and the more general provisions. It is our understanding that the tech industry has developed, and is developing systems to improve tackling this harm. We feel this is not adequately explored or addressed by the consultation.

We are also unclear as to how collector sites - that exist for the purpose of file / portfolio sharing of image based sexual abuse material fit within these requirements, and extent they will be 'caught' by the requirements. Our understanding is that they would not be.

Scope

36. As with other measures, these measures won't apply to small but risky sites, which are often specifically used to share non-consensual images between users. We also suggest that the measure should apply to deepfakes as well as "real" intimate images, as deepfake porn can have devastating effects too. It is suggestive not just of intrusion, but of a deliberate choice not to seek consent.

Do you have any evidence on the relative efficacy of third-party and internal databases for image-based IIA content?

37. There is a risk of providers deciding to use other third-party services beyond StopNCII, which has a strong track record of effective survivor-centred take down, and survivors having to report multiple times through multiple different platforms. There is no standard set for third-party services - if a provider of a hash-database has a poor track record then the database they are using is likely to be inadequate, or inferior to StopNCII's. Ideally there would be a central coordination measure or a central database so hashmatching would be shared via different services and there would be improvements across platforms.

38. The approach to hashmatching for IIA is modelled on the successful work of StopNCII whose work has been groundbreaking in this area. While Ofcom has recognised that different providers may emerge, there is a risk that this results in an uncoordinated approach which could lead to significant inefficiencies or re-traumatisation of survivors having to submit intimate images of themselves to multiple providers. Ofcom should consider how best to mitigate this risk, perhaps through strong recommendations favouring interoperable systems, setting standards for an acceptable database and/or by building on the existing trusted model of StopNCII as the market leader.

Perceptual hash matching for terrorist content

Do you agree with our proposals? Please provide your reasoning, and if possible, provide supporting evidence.

39. We broadly support these proposals, which will provide consistency with CSAM and NCII by introducing further duties for platforms to use hash matching for terrorist content.

40. However, in support of points raised by Tech Against Terrorism in their submission, we are concerned by Ofcom's reluctance to use third-party providers of hash lists, which they state is due to the fact that third-party providers don't necessarily align with the definition of terrorist content as provided by the OSA. Whilst it is true that third-party providers do not curate their collections of terrorist content, which they hash, to suit the UK's statutory definitions, Tech Against Terrorism believe that the time, money and effort required for platforms to build their own lists would be colossal and insufficient. This is because in-house teams, if they can be spared from general duties, will not be engaging in the kind of cross-platform monitoring required to maintain the breadth and depth of coverage necessary to block terrorist content at the point of upload.

41. The element of human review which will inevitably be required to review positive matches should suffice to cater to different jurisdictional requirements, and the largest platforms, who are likely to be within scope of the measures, are well-versed in doing just this. In Germany, for example, content denying the Holocaust is banned, and this ban is often observed in Germany but nowhere else. Such differentiated approaches to compliance are easily practicable.

Perceptual hash matching for child sexual abuse content

Do you agree with our proposals? Please provide your reasoning, and if possible, provide supporting evidence.

42. We welcome Ofcom's proposals on CSAM hash-matching, particularly the extension of hash-matching duties to apply to high-risk porn providers regardless of their size, which will be vital in taking down CSAM and acknowledges arguments we have previously made about small but risky sites. We are also pleased to see that there is a wider range of CSAM in scope for proactive detection.

43. As we have addressed in much further detail our analysis on technical feasibility, we are also concerned that Ofcom will be undercutting the positive proactive-tech measures that they have proposed with the technically feasible clause. Whilst Ofcom have said they will ask platforms to keep a record of why a service cannot implement the proactive tech for them to review, however

it isn't yet clear how Ofcom will investigate services' claim that there is no technology which meets the criteria.

44. We defer to the expertise of the Internet Watch Foundation who have set out more detail in their response about their concern regarding Ofcom's proposed criteria for proactive technology, which is designed with AI detection tools in mind rather than hash-matching.

Recommender systems

Do you agree with our proposals? Please provide your reasoning, and if possible, provide supporting evidence.

45. There is a wealth of research that demonstrates the way in which social media platforms, such as TikTok[10] and YouTube[11], profit from harmful recommender settings driven by algorithms. Indeed, Ofcom acknowledges lessons learnt from the Southport riots, where harmful misinformation about the immigration status of the perpetrator spread rapidly before verification.

46. We welcome the proposed measures for recommender settings that would bring the Illegal Harms Code in line with existing recommendations in the Protection of Children Code. However the changes do not represent a radical shift in approach to how recommender systems are regulated. The safety-by-design approach Ofcom has taken is ex-post, relying on automated functionality being built into systems to pause amplification, including real-time filtering, ranking adjustments and delay mechanisms within recommender algorithms. This means that, much like the principles in the Protection of Children Code, the measures are about content rather than systems and service redesign, despite the recognition of design choice and the importance of the relevant weight attached to different signals (para 14.6) and the specific risk of optimising for engagement (paras 14.7, 14.9).

47. Ofcom have explained in stakeholder sessions that this functionality is expected to be built into systems so that its operation is almost automatic, rather than switched on or introduced manually ex post. This must be further elucidated in the final guidance to avoid ambiguity.

48. We also recommend that CSAM and NCII are included. Recommender tools were part of the spread of the deepfake images of Taylor Swift.[12]

[10] https://counterhate.com/research/deadly-by-design/

[11] https://www.tortoisemedia.com/2024/12/11/youtubes-algorithm-recommends-eating-disorder-content-to-teenage-girls

[12] https://www.huffingtonpost.co.uk/entry/the-issue-of-sexually-explicit-deepfakes-is-far-larger-than-taylor-swift_uk_65cdda55e4b0dd11b911faab?origin=related-recirc

49. It should also be extended to accounts that sell or promote CSAM, even if they do not directly share CSAM. There is evidence that these accounts have been promoted on Instagram through recommender systems. Although some of the measures in this consultation and in the illegal harms codes will mitigate these risks (user banning and CSAM URL sharing), including these accounts in these measures will help strengthen protections.

50. Whilst we support limitations on the reach of content that is harmful in nature, there is a need for strong user transparency safeguards to explain how recommender systems will work in practice and notify creators when their content is being held back or deprioritised under this measure so as to allow them to use the complaints and appeals processes envisaged by the Act - and an important part of the safeguards around freedom of expression.

Do you agree with our assessment of the impacts (including costs) associated with this proposal? please provide any relevant evidence which supports your position

51. We note and support Ofcom's assessment of the impact of the measures on freedom of expression and agree that they are likely to be proportionate and also recognise the benefits to freedom of expression outlined in para 14.59. We reiterate our comments about transparency, and emphasise the need for speedy moderation practices: to ensure that this is not at the expense of accuracy there is a need for adequate resourcing. While we agree that there are costs in dealing with user complaints (noted in para 14.52), we are not convinced that they are such as to correct for any risk of under-resourcing. Ofcom should make this clear to services, especially since the Codes give them latitude in how they resource content moderation. Please also see our comments on the scope of the ICJG. Should NCII and CSAM be included in the measure, there would also be the added impact of protecting Article 8 and Article 3 rights.

User-banning and preventing return following detection of child sexual exploitation and abuse content

Do you agree with our proposals? Please provide your reasoning, and if possible, provide supporting evidence.

52. We welcome proposals to introduce user bans for adult users who share, generate or upload CSEA, which can be an effective way of tackling repeat offenders and ensuring that children are not put at harm by someone who is already known to a platform. We would urge Ofcom to consider widening the scope of these measures to include NCII in order to bring them in line with Ofcom's hashmatching proposals.

53. However it is important for Ofcom to take a safety-by-design approach, rather than relying on user sanctions. User bans should be accompanied by preventative measures such as warning messages and restricting adult users ability to interact with child user accounts. This must also be supported by educational measures to reduce reoffending, including signposting to resources on platforms.

54. There is also a concern that adult users committing harm against children who have been banned from larger user-to-user sites will be pushed onto smaller platforms, where Ofcom has advised against more stringent measures. This is not an argument for reducing the measures proposed but for extending them to small but risky services.

55. Currently, there is no requirement for sites to share information with each other about users that have been banned from their platform. Including such a requirement would allow platforms to proactively combat further instances of CSEA from occurring.

56. As we will discuss in more detail in the next question, Ofcom must take a more nuanced approach to banning accounts of child users.

What is your assessment of the options we set out in relation to the treatment of child users and which option do you consider to be most appropriate? Please provide any supporting evidence to support your arguments.

57. The treatment of child users within the proposed user banning framework should be guided by a nuanced understanding of both risk and responsibility. Children occupy a unique position online, they can be victims, bystanders, and sometimes perpetrators of harm, often without full awareness of the consequences of their actions. Both permanently banning or entirely exempting children from sanctions would fail to reflect this complexity. A sliding scale which takes into account the different contexts (discussed below) may be more appropriate.

58. Ofcom should aim to protect children who have been victims of abuse, whilst also recognising that punitive measures against children who have perpetrated abuse may in some circumstances exacerbate the harm. This should include children who have been coerced into sending images that are classed as CSAM, such as self-generated CSAM. Permanent exclusions from online spaces risk isolating vulnerable young people by discouraging them from reporting incidents or pushing them towards less safe, unregulated services.

59. Ofcom already recognises that this requires a principles-based approach in the consultation document, which would ensure that action is taken on reports of child-on-child CSEA in a proportionate manner. Sanctions should therefore follow a sliding scale that recognises the many variables, including the age, maturity stage and motivation behind the incident. This should include educational interventions and support alongside stronger measures where the behaviour is deliberate, repeated, and poses an ongoing risk.

Do you agree with our assessment of the impacts (including costs) associated with this proposal? Please provide any relevant evidence which supports your position.

60. We agree that the severity of CSEA justifies the measures taken in free speech terms. We would also note that for adults sharing children's images, their speech is of a low order (if not excluded altogether by Article 17 from Article 10), attracting less protection. Children's sexual speech is more difficult to assess as they may be expressing an essential part of their personality, which should be relatively highly protected under Article 8, or they may be punishing an ex, or bullying a peer.

Highly Effective Age Assurance (HEAA)

Do you agree with our proposals? Please provide your reasoning, and if possible, provide supporting evidence.

Defining HEAA

61. We agree with and welcome Ofcom's decision to extend highly effective age assurance across the Illegal Content Codes and align the presentation of HEAA in the Illegal Codes with the Protection of Children Codes, closing an important gap in provisions. In particular, providing more clarity around the definition of HEAA, including requirements on providers to consider usability, fairness and privacy when determining the age of a user. In an evidence session with the Communications and Digital Committee, Baroness Kidron noted that unless age checks are "radically private" they will be culturally rejected, emphasising the point that data gathered on a user must be limited strictly to age[13].

Applying measures to all users

62. Turning HEAA on for all users risks flattening the ability to deliver child-specific messaging or support in grooming contexts. We therefore urge Ofcom to ensure that the age assurance section explicitly advocates for a hybrid approach which would introduce a baseline mandatory HEAA across all users, but with adaptive, more protective layers triggered in contexts of higher risk, such as with livestreaming or settings where there is a higher risk of grooming, and with distinct, prioritised support pathways for users identified as children.

63. We noted in our response to the Protection of Children Codes that Ofcom has not attempted to introduce measures that would take into consideration the different age groups of children who might be on platforms and how harm manifests itself according to age. If platforms know the age

[13] https://committees.parliament.uk/oralevidence/16484/pdf/

of their users, it should be possible for them to introduce different measures for those different users.

Backdating/ alignment

64. Ofcom must ensure there is alignment between HEAA requirements and the Age Appropriate Design Code (AADC). While Ofcom's proposals set out a definition of HEAA in the Protection of Children Code, there remains uncertainty about how HEAA obligations will interact with other safety measures where such duties are not explicitly mandated. Embedding AADC principles across the application of HEAA would provide stronger regulatory coherence.

65. Safety and data protection cannot be a trade-off. HEAA measures must not conflict with GDPR requirements if age-assurance data is repurposed or handled without strict privacy safeguards. While Ofcom recognises this point, alignment with AADC would address these concerns by clarifying that HEAA must always be privacy-preserving and purpose-limited. There is also scope for Ofcom to be more ambitious in defining who falls within scope of these measures because a wider range of services are already subject to the AADC. Adopting HEAA as a default requirement would simplify compliance by extending protection consistently across platforms.

Do you agree with our proposal to introduce age assessment appeals measures into the Illegal Content User-to-user Codes (ICU D15 and D16)? Please explain your reasoning.

66. We agree with introducing age assessment appeals but highlight that this can introduce perverse incentives for services to overestimate user ages (as complaints will come from adults appealing underestimation). Therefore, we recommend Ofcom introduces a minimum expected accuracy of age assurance mechanisms (e.g. 95% as recommended by AVPA) to ensure technical accuracy remains high.

Increasing effectiveness for U2U settings, functionalities and user support

Do you agree with our proposals? Please provide your reasoning, and if possible, provide supporting evidence.

67. We welcome Ofcom's proposal to require providers to either implement HEEA to apply ICU F1 (default safety settings for child user accounts) and ICU F2 (supportive messaging features) to all users that have not been determined to be an adult, or apply the measures to all users. Giving providers flexibility between the two options is reasonable, but in both cases children will benefit from stronger safeguards against unwanted contact.

Crisis Response

Do you agree with our proposals? Please provide your reasoning, and if possible, provide supporting evidence.

68. We welcome action from Ofcom to learn from events such as the Southport Riots, which saw online misinformation turn into offline acts of violence, by introducing an obligation on platforms to create a Crisis Response Protocol (CRP) to detect crises and provide a dedicated channel for law enforcement, as well as a post-crisis analysis to review and evaluate the platform response when requested by Ofcom. We agree with the points made in para 20.42-48.

69. We recognise that having processes in place to identify crises and appropriate responses are an essential starting point in this space, however the measures lack clear operational expectations or mechanisms for cross-platform coordination. Cross-platform coordination is a significant gap given the cross-platform nature of crises such as the Southport riots; search engines may also have a part to play and, while we do not have evidence to submit on that in relation to Southport, we would recommend Ofcom considers any available evidence or expert insight from other organisations on whether protocols should be applied to those services as well.

70. The focus on monitoring and evaluation through crisis-analysis is welcome, but while Ofcom requires providers to keep a record of their analysis it does not require them to publish it or share it routinely with Ofcom. There should be greater public transparency about their response to a crisis.

71. We are concerned that this measure does not apply to public health crises. While we understand that the imperative for introducing the measure was in response to the post-Southport riots and Ofcom needs to take account of the alignment with the illegal harms duties, the way in which public health crises can evolve online is similar to the type of information incidents that can lead to violent disorder and civil unrest. It feels like a wasted opportunity for Ofcom not to include this for consideration when the governance measures it is proposing would be equivalent across these types of crises. It also suggests that Ofcom requires a public health crisis to have taken place before it will consider that that constitutes a crisis that should be prepared for (as is the case with the introduction of this measure after the fact of the post-Southport violence and civil unrest). Waiting for evidence to act risks putting many more people at risk of harm than preempting the potential need to prevent that harm in the first place.

Do you agree with our proposed definition of 'crisis'? Please explain your reasoning, and if possible, provide supporting evidence.

72. Ofcom's definition of crisis focuses on public safety. It does not, by contrast to the position under the Digital Services Act (as noted in para 20.29) or the "special circumstances identified in

s 175 OSA, refer to public health. While we appreciate that Ofcom may want to focus this measure on acute rather than chronic problems, there may be some circumstances where there is a public health crisis that is not well encompassed by the idea of public safety. There does not seem to be a justification for this exclusion; we suggest that public health be added as an additional component. Given the requirement also for there to be an "extraordinary situation" there are natural limitations to the circumstances in which the protocols will be triggered. This may be especially relevant when we consider content harmful to children, though this would take us beyond the types of priority content identified in para 20.27, to include content listed at s 62(9) OSA.

73. We assume that a crisis can be one taking place on a particular service, but also a crisis in real life; it would be beneficial for Ofcom to reflect this in its guidance, as well as the fact that a crisis need not be national - it can be regional or local. The examples given in para 20.29 imply both these points, but to have the point made expressly would be useful in understanding when the CRPs should come into play (and thereby be useful in enforcement terms).

74. The measure uses a limited, three-stage definition of a crisis rather than recognising the full life cycle of a crisis that may increase in intensity and require more nuanced staged responses. The Government recognises tiered threat levels, with different levels of intensity of response. Moreover Full Fact has developed a five stage model which identifies how services can pick up signals of impending crises. While this does not require action about sub-criminal content (that is not harmful to children), it should indicate that services should have regard to it. Identifying the end of a crisis is just as important. this may especially be the case where an acute situation remains on-going. At what point might a service provider be entitled to consider that such a situation is the new normal. Further guidance on this would be helpful.

75. There is also a concern that Ofcom leaves individual platforms to identify their own indicators for monitoring and identifying a crisis. Whilst several examples are given, such as 'law enforcement', platforms would have more clarity if these were expanded. It is also unclear whether Ofcom will verify the appropriateness of the indicators chosen, or whether the fact that there are some indicators alone would suffice. Platforms' mechanism for triggering the crisis protocol needs to be effective. Standards that guide this, as well as how fast the crisis protocol is enacted, would provide more clarity for platforms. This would also aid platforms to have a more consistent approach, which is particularly important given the lack of clear guidelines for cross-platform responses. It also would provide some protection against any particular weaknesses in moderation function (eg through under-resourcing) undermining this measure, especially as it is anticipated that crisis response could rely on the re-allocation of existing resource (see para 20.36) not the deployment of extra.

76. We would also question whether Ofcom should be making their proposals for post-crisis analysis (20.38 and 20.50) stronger. In the first reference, Ofcom proposes providers "should" conduct one and in the second, Ofcom notes "a post-crisis analysis should drive improvement in

providers' systems and processes for dealing with a crisis and identify gaps within the provider's wider trust and safety systems and processes." Given the necessity for specific measures elsewhere in the codes and the safe harbour provision, this seems less robust: should it not be the case that providers "must" conduct a post-crisis analysis and "must" use the learning from it to drive improvements, with Ofcom assessing their compliance on those terms?

Is there any evidence of best practice in responding to a crisis that we have not identified? Please explain your reasoning, and if possible, provide supporting evidence.

77. We note in 20.32 that Ofcom has "evidence to show that some service providers already have some form of crisis response mechanisms in place". It is not clear whether Ofcom has asked for further information from these services as to their effectiveness or whether the existing mechanisms form the baseline of good practice on which this code measure should build. Given the emphasis on consultation respondents providing evidence to help the regulator in its work, if Ofcom has not done so, this would seem like an oversight. Just having a crisis response mechanism does not equate to it being effective in a crisis, just as mandating such a mechanism in a code does not equate to services taking effective action when a crisis occurs. With the evidence already available, Ofcom might have been expected to be able to provide some additional requirements for this iteration of the measures.