Protecting Children from Harms Online
OFCOM Consultation

response from

Center for Countering Digital Hate (CCDH)
July 2024

Introduction

About CCDH

The Center for Countering Digital Hate (CCDH) is a non-profit working to stop the spread of online hate and disinformation through innovative research, public campaigns and policy advocacy. From offices in London and Washington DC, the Center works to protect human rights and civil liberties online, hold technology companies responsible for their business choices, and advocate change to protect our communities.

Structure of CCDH's submission

The submission opens with our overarching analysis of the materials drafted by Ofcom in this consultation. CCDH then responds to specific questions relevant to our work from the consultation form. All CCDH research reports linked or cited in this consultation follow as appendices.

Answers to questions 9 and 43 are marked confidential, as they cite evidence CCDH has not yet published.

Overarching analysis of consultation and draft materials

CCDH has supported the Online Safety Act since its inception. We commend Ofcom for the extensive work put into drafting the materials under review in the "Protecting children from harms online" consultation. We feel these will go a long way to improving online safety for children in the UK.

As guided by our global standard for regulating social media, the CCDH STAR Framework, CCDH believes these proposals can be strengthened and offers evidence from our research to support the additional obligations and measures we suggest.

We understand that Ofcom has not incorporated the feedback it received in the illegal harms consultation into the draft proposals here (as stated in Vol. 1, 3.3). CCDH therefore repeats some points made in the earlier illegal harms consultation response.

Overall, our concerns with the proposals are that they (1) do not meet the safety by design imperative set out in section 1 of the Online Safety Act; (2) are overreliant on the iterative nature of the Codes; and (3) do not adequately translate the risks identified into proposed measures.

(1) Safety by design – The Online Safety Act states that the duties imposed by the legislation seek to ensure that regulated services are "safe by design" and operated in such a way that delivers a higher standard of protection for children than for adults (Section 1(3)). The UK Government defined 'safety by design' in guidance on the principles of safer online platform design as: "the process of designing an online platform to reduce the risk of harm to those who use it. Safety by design is preventative. It considers user safety throughout the development of a service, rather than in response to harms that have occurred". However, the draft codes largely concern content and contain limited proactive measures. CCDH believes that a regulatory approach with safety by design at its core would oblige service providers to consider outcomes *in addition* to content moderation and mitigation measures.

(2) Iterative nature - the draft codes of practice are first iterations, but CCDH believes Ofcom runs the risk of setting a low bar at the outset of the regime. Until we see evidence to the contrary (for instance, in Ofcom's response to the illegal harms consultation), we are concerned that the proposed framework will not be substantially improved in subsequent versions of the codes and that the focus on content-moderation and tick-box compliance will become the baseline for the regime going forward.

(3) Translating identified risks into proposed measures – many of the risks identified in Ofcom's evidence base (Volume 3) do not have corresponding measures or required outcomes in Volume 5. CCDH examines these in response to question 22 below. Ofcom should ensure proposed measures address all identified risks. CCDH will advocate that this be a condition of the safe harbour protection the Act grants (Sections 11, 28) to services which adhere to the codes of practice.

Volume 2: Identifying the services children are using

1. Do you agree with our proposals in relation to children's access assessments?

Introducing age gating - as mandated by the Act - is a first and critical step towards keeping children safe online by preventing them from using products or services evidence shows are harmful to them. Ofcom's proposals in Volume 2, to identify the services children are using, are the entry point for delivering that mandate.

In Volume 2, the age gating requirements following the children's access assessment sit atop all the other obligations on regulated services. The application of safety technologies on top of a system deemed harmful to users might leave unaltered the online ecosystem beneath it. CCDH is concerned that this sets age gating up as the critical point in the regulatory regime, doing all the heavy lifting to achieve a regime which has higher safety requirements for children than for adults. As such, age gating risks becoming a single point of failure that undermines the child safety regime.

In addition to the measures relating to recommender systems (RS1, RS2), age gating is the only new measure to protect children beyond measures pulled across from the duties on illegal harms. The risk assessment obligations for children appear no more stringent than the proposals in the illegal harms consultation, nor does CCDH see a requirement here for any significant redesign of service following any

risks identified in the children's access risk assessments ([A6](#)). Choosing simply to keep children off a regulated platform appears, to us, to dilute the safety by design imperative set out in section 1 of the Act to design and operate safer services.

Although age assurance does not always satisfy a "safety by design" assessment, being fundamentally a means of keeping children off unsafe platforms rather than incentivising safer platform design for all users, CCDH does note and commend Vol 2 for moving away from prescriptive tick-boxing measures such as those we flagged in our [illegal harms consultation response.](#) For example, in Vol. 2, 4.55, Ofcom reflects that it is *for services* to understand the effectiveness of their age assurance measures, access control methods, and processes through their own testing and enquiries of age assurance providers. Critically, it instructs services that if evidence suggests any reduction in the efficacy of their measures, that they must repeat the children's access assessment without intervention from the regulator. CCDH commends the onus of responsibility here.

There is also an example of the outcomes-based requirements CCDH is advocating for in this volume: the proposed criteria for assessing the effectiveness of age verification (technical accuracy, robustness, reliability and fairness) are more about outcomes than specific outputs (Vol. 2, 4.9). CCDH proposes that analogous criteria are introduced to assess the processes adopted to identify harms and select appropriate mitigation measures in later volumes.

Volume 3: The causes and impacts of online harm to children

7. Do you have any views on our interpretation of non-designated content or our approach to identifying non-designated content? Please provide evidence to support your answer.

16 Non-designated content (NDC) is a distinct category of content under the Online Safety Act. It is defined as content that is not PPC or PC, but which presents a material risk of significant harm to an "appreciable number of children in the United Kingdom". In preliminary assessment, Ofcom identifies two broad kinds of content that could meet the definition of NDC: "body image content" and "depressive content". "Body image content" can lead to harm arising from body or image dissatisfaction, including low self-esteem, impacted mental health, and harmful behaviours such as extreme dieting and exercise. "Depressive content" can also have lasting emotional impacts, including exacerbation of mental health issues such as depression, anxiety, self-harm and suicidal ideation.

CCDH agrees with Ofcom that aside from being considered categories of NDC, body image content and depressive content already fall within scope of the regulation via the context of cumulative harm (Volume 3, 7.9.8). But our research, below, supports the proposal to include these as specified NDC categories.

8. Do you have any specific evidence relevant to our assessment of body image content and depressive content as kinds of non-designated content? Specifically, we are interested in examples of body image or depressive content linked to significant harms to children.

18 CCDH continues to research the prevalence and impact of harmful body image content on social media platforms. Our research on the promotion of steroid-like drugs has application in many areas of Ofcom's online safety regulation, but should also be considered under NDC body image content. By promoting a body ideal achieved through the use of steroid-like drugs, this body image content increases the likelihood of children seeking out and using those substances -- substances linked to a significant number of harmful, even fatal consequences.

19 In our 2023 report *TikTok's Toxic Trade*, CCDH evidenced the widespread promotion of steroid-like drugs on TikTok. While these drugs are illegal for human consumption (UK Food Standards Agency) they exist in a legal grey area in which they are permissible for sale as research chemicals. Social media platforms' policies on the promotion and sale of illegal drugs rarely take account for such ambiguity, and thus the content continues to circulate largely unchecked. While CCDH has consistently argued that content featuring steroid-like drugs should be recognised and moderated under platforms' illegal drug policies, CCDH will leave the legal aspects of this content aside to assess steroid-like drug content within the parameters of "body image content" described in Vol.3, 7.9.

20 CCDH found that UK TikTok users viewed content posted to steroid hashtags on TikTok 117 million times from September 2020 through September 2023. UK TikTok users aged 18-24, the youngest age range CCDH researchers could analyse using TikTok's Creator Center analysis tools, accounted for 89 million of those views. Although TikTok does not make viewership data available for under-18s, given the high number of underage users Ofcom knows to be using TikTok, an appreciable number of children in the UK are likely to have been exposed to this content on their For You Page and via TikTok's recommender system.

21 CCDH found that content posted to these hashtags promoted the use of steroids and similar drugs while downplaying the potential risks. Some videos using these hashtags encourage viewers to "just tell your parents they're vitamins" or advocate for viewers to start using steroids and steroid-like drugs in their teens.

22 To be clear, the body image promoted by this content goes beyond 'fitness content' that would not count as NDC. This content promotes a body image that requires the procurement of substances illegal for human consumption and has intense physiological consequences, especially on children. Even if viewers do not go so far as using the steroid-like drugs featured in this content, the promotion of bodies altered by steroid-like substances will have psychological effects on viewers: normalising this as a desirable body and increasing body dissatisfaction by comparison. The mental health consequences are sometimes referred to as 'bigorexia'.

23 There are appreciable harms to physical health too. Research on steroid-like substances from the US Food and Drug Administration evidence the consequences for the body including: increased risk of heart attack or stroke; psychosis and hallucinations; sleep disturbances; sexual dysfunction; liver injury and acute liver failure; infertility and sexual organ disfunction. The risks, particularly of irreversible consequences on sexual functions, are increased if steroid-like substances are consumed by children who have not finished the body's natural maturation through puberty.

24 Content that promotes or glorifies the use of steroid-like drugs poses the risk of significant harm to an appreciable number of children in the UK. CCDH believes such content meets the definition of NDC under the Act and that platforms should therefore make steps to protect children from encountering it online.

9. Have you identified risks to children from GenAI content or applications on U2U or Search services?

✂

Volume 4: How should services assess the risk of online harms?

15. Do you agree with the proposed measures included in the Children's Safety Codes? If you responded to our Illegal Harms Consultation and this is relevant to your response here, please signpost to the relevant parts of your prior response.

In our illegal harms consultation response, CCDH raised concern that the prescriptive approach to the codes risked them becoming a tick-box exercise for regulated services. Specifically, that the safe harbour offered by the Act for services who comply with measures in the codes could lead to services dropping protective measures beyond those prescribed, or ceasing application of innovative new ones. CCDH thus commends the clarification on this point contained in Vol. 4, Table 12.2: that "stopping implementing such measures or changing them may constitute a significant change (see Step 4 below) and may increase their risk level". CCDH hopes that this reduces the likelihood that regulated services stop applying existing protective measures not explicitly required or recognised as compliant with the regulations.

Volume 5: What should services do to mitigate the risk of online harms

22. Do you agree with our proposed package of measures for the first Children's Safety Codes? If not, please explain why.

32 Functionalities and features identified in the Volume 3 register of risks do not have corresponding mitigation measures in Volume 5. This is the same problem as was identified in the earlier illegal harms consultation.

As remarked upon above, CCDH commends Volume 3 for the thorough and detailed analysis of the causes and impacts of online harms to children, but feels that this work is insufficiently translated into mitigation measures laid out in the codes of practice for user-to-user services (annex 7) and search services (annex 8). These measures focus on content takedown and downstream measures to deal with PPC and PC content already circulating on a service (the superior recommender system measures are analysed later in response to questions 49 and 50).

CCDH points to more extensive work on this topic from the Online Safety Act Network, particularly on the gaps in Volume 5 around livestreaming. From our own work, CCDH is concerned about a lack of measures dealing with hashtags and keyword tags, which are correctly identified in Volume 3 as functionalities with significant risks to children in terms of exposing them PPC and PC content. Our research on this topic fits within both the illegal harms and protecting children areas of online safety regulation, but is submitted here.

CCDH believes there is sufficient evidence to include measures requiring services to assess and mitigate the risk arising from hashtags known to be associated with PPC, PC, and ND content. This will be an iterative process as language evolves and users adapt to avoid moderation.

In CCDH's 2022 report *Deadly by Design*, researchers used hashtags to identifying features by which recommender systems promoted different types of content to the accounts of 13-year-old girls. The hashtags themselves seemed to play a facilitatory role in bringing vulnerable children into online communities promoting harmful behaviours, such as self-harm. CCDH found that hashtags on eating disorder content had over 13.2 billion views.

In current research, CCDH is examining pro-eating disorder hashtags and keywords on YouTube. Our preliminary findings show that pro-eating disorder terms are being utilised to signpost videos to vulnerable individuals seeking eating disorder content. These terms are searchable via the service's search function. This is despite YouTube's policies stating that it doesn't host harmful eating disorder content.

Examples on YouTube include instructions on how to engage in eating disorders (including how to conceal them) and content about eating disorders that features imitable behaviour including disordered eating behaviours, such as purging after eating or severely restricting calories. Weight-based bullying in the context of eating disorders is also included. CCDH will submit this research to Ofcom once published.

It is clear from CCDH's research that hashtags and keyword tags are vectors in the discovery and promotion of content harmful to children. On the basis of this evidence, CCDH argues for explicit proposals added to Volume 5 and the Children's Safety Codes to address content associated with hashtags and keywords that should not be recommended to users.

27. Do you agree that most measures should apply to services that are either large services or smaller services that present a medium or high level of risk to children?

No. While accepting that the Act establishes differing responsibilities based on service size, aspects beyond the scope of this consultation, CCDH believes that safety by design should be required of all online services, regardless of size. Within the scope of this consultation however, there are specific measures dealing with safety by design, such as written statements of responsibilities (Vol. 4, GA3) and expectations of product testing (Vol. 4, Table 12.3), in which Ofcom appears to have made a judgement that it is only reasonable to expect large or multi-risk services to comply with the safety by design requirements. CCDH believes this is a judgement which should be reconsidered, and within scope of Ofcom's consultation, revised to apply more broadly.

28. Do you agree with our definition of 'large' and with how we apply this in our recommendations?

It appears to CCDH that the size bands proposed in the Illegal Harms Code (Vol. 3, 9.98c) are the basis of the child specific bands in the Children's Code. This presumably means that only services that have more than seven million monthly child users must comply with many of the measures unless they also meet the multi-risk requirement. The number of services meeting this threshold is miniscule compared to the number of services in scope of the Act. A service should not be able to accrue seven million monthly users before it must prove itself safe for children. Given the evidence of the extreme harm from small, high-risk services submitted to Ofcom in the Illegal Harms Code of Practice consultation responses, CCDH is concerned with this recommendation.

43. Do you agree with the proposed user reporting measures to be included in the draft Children's Safety Codes?

✂

49. Do you agree with the proposed recommender systems measures to be included in the Children's Safety Codes?

CCDH feels that the new measures relating to recommender systems (RS1, RS2, RS3) are significant steps forward in increasing the protections for children, particularly in relation to reducing their exposure to Primary Priority Content and Priority Content. But the limitations of the measures in addressing wider safety by design factors remain. For example, many of the measures proposed to address recommender systems are downstream, implemented as mitigation measures long past the system design and product development stages. CCDH reiterates that a safety by design process, in which there is rigorous product safety testing and risk assessment, would require measures earlier in the process.

50. Are there any intervention points in the design of recommender systems that we have not considered here that could effectively prevent children from being recommended primary priority content and protect children from encountering priority and non-designated content?

Yes. Earlier intervention points increase the likelihood that measures will lead to safety by design outcomes and effectively prevent children from encountering PPC and PC. Specifically, CCDH believes there is a gap in the analysis of recommender systems relating to the risks arising from the business model that manifest as harms to children *via* the recommender system.

Ofcom's description of recommender systems risk in Vol. 3 states that the technical functions of recommender systems are not, in themselves, harmful: "the functionalities and characteristics we describe as risky are not inherently harmful and can have important benefits. For example, recommender systems benefit internet users by helping them find content which is interesting and relevant to them. The role of the new online safety regime is not to restrict or prohibit the use of such functionalities or characteristics, but rather to get services to put in place safeguards which allow users to enjoy the benefits they bring, while managing the risks appropriately." (Vol. 3, pg 4).

The use of "safeguards" here, by CCDH's reading, implies that recommender systems can run as previously but must now implement interventions designed to meet the downstream measures described in Vol. 5. This fails to implement safety by design principles.

55 CDH believes further work can investigate the business incentives which underpin recommender systems, the values they incorporate, and ways those two forces together propagate harms to children. The consultation also does not consider how recommender systems form part of the suite of incentives for content creation, engagement, and how being picked up by the algorithm is important for users' advertising revenue and promotion.

56 CCDH proposes that Ofcom flesh out the relationship between the disparate business model risks and recommender system risks currently contained in the Vol. 3 register. From this evidence, it is then possible to propose safety by design measures at earlier intervention points in Ofcom's proposals for application to recommender systems.

Closing remark

CCDH is grateful for the opportunity to provide evidence to Ofcom's protecting children from harms online consultation. We will continue to be constructive partners and support the regulator in work to implement the world-leading Online Safety Act regulatory regime in the UK.