WESTMINSTER HALL: IMPLEMENTATION OF THE ONLINE SAFETY ACT

Joint Civil Society Briefing

9.30am-11am, Wednesday 26th February 2025

5RIGHTS FOUNDATION
CCDH
NSPCC
ONLINE SAFETY ACT NETWORK
Clean up the internet
IWF Internet Watch Foundation
MOLLY ROSE FOUNDATION
internet matters.org
LUCY FAITHFULL FOUNDATION
Thomas William Parfett Foundation
global action plan

The Online Safety Act places strict requirements on tech companies to make sure their platforms are not harming children, or being used to harm children. However, Ofcom's current codes of practice will not meet this duty and will leave children exposed to harm. This is contrary to the purposes of the Act and the intention of Parliament.

We are asking MPs to call on the Government and Ofcom to take action so that the codes of practice:

1. Require services to **take a safety by design approach**, reflecting the overarching duty within the Act and ensuring its codes address all known risks to children.

2. Revise its approach regarding prescriptive safety measures to ensure that extensive evidence from civil society and victims' groups and the cost to victims are acted upon to **reflect the Act's risk-based regulatory regime**.

3. Require services to **give separate consideration to children in different age groups** by enforcing age-appropriate design.

Action **must** be taken ahead of the first Children's Safety Code, due April, and the next iteration of the Illegal Harms Code to ensure robust protections for children.

INTRODUCTION

The Online Safety Act (2023) places strict requirements on tech companies to assess and mitigate the risk of foreseeable harm on their services – in particular illegal harms and harms with specific risks to children. This was the overarching intent of the legislation when it passed through Parliament in October 2023.

The UK's online safety regulator, Ofcom, has been tasked with implementing the regime by providing detailed guidance and codes of practice and setting out how it will approach enforcement. In December, Ofcom published the first iteration of its Illegal Harms Code of Practice[1] and is due to publish its final Children's Safety Code of Practice in April 2025.

Since the publication of the first draft code, there has been considerable concern from civil society, victims' groups, academics, parents' groups and parliamentarians that Ofcom's proposals do not align with parliamentary intent and will fail to meet the Act's objectives.

[1] Statement: Protecting people from illegal harms online

The primary reasons for this shortfall are twofold:

- Ofcom has adopted an overly narrow legal interpretation of the Act, which limits its ability to recommend and enforce robust safety measures.
- The codes set a low baseline for compliance, failing to establish sufficiently high safety expectations for regulated services.

Despite extensive evidence from civil society[2] on where the codes could go further, Ofcom has chosen to not substantially change its approach.

This briefing sets out some of the most pressing issues with implementation, measuring Ofcom's interpretation of the Act against its parliamentary intent, and signals why action is needed to ensure the first Children's Safety Code, and the next Illegal Harms Code, are substantially more robust.

1. SAFETY BY DESIGN

"Safety by design" is the principle that companies ensure known or anticipated harms have been evaluated and addressed in a service's design. This means that companies should assess and mitigate risks before they occur by implementing preemptive safety measures. Safety by design practices are crucial for keeping children protected on online services but are not currently meaningfully adopted by tech companies.[3]

Section 1 of the Act establishes that the purpose of the Act and its overarching duties for tech companies are that **services will be made safe by design** and **provide a higher standard of protection for children than adults**. Safety by design includes all aspects of the services – including its features and functionalities, such as livestreaming and private messaging.

While Ofcom's register of risks acknowledges the role of a range of functionalities, its codes do not reflect this and are almost exclusively concerned with content. This is counter to the stated aims of the Government who said at the time: "… This is not just a content Bill… The ways in which a service is designed and operated, including its features and functionalities, can have a significant impact on the risk of harm to a user."[4]

Ofcom's proposals would also not address risk upstream, with most of its suggested compliance measures applying only after a risk is already active – including content moderation, reporting and complaints. This exposes children to harm, overemphasises content issues, and downgrades issues relating to design and systemic risk.

What this means in practice:

The lack of consideration of features and functionalities runs contrary to duties in the Act that explicitly state they apply "across all areas of a service" and require services to take measures relating to "the design of functionalities, algorithms and other features."[5]

In addition, several features are not addressed in Ofcom's measures, despite the evidence of the ways they facilitate harm. This includes:

[2] See: Online Safety Act Network, Resources (Consultation responses)

[3] Online Safety Act Network, Safety by Design

[4] Lord Parkinson of Whitely Bay (19th July 2023) Online Safety Bill, Report Stage (5th Day), col. 2419

[5] s.12(8), Online Safety Act 2023

- Livestreaming;
- Content available for a limited time (ephemeral content);
- Recommender systems designed for extensive use (including entrapping users in 'filter bubbles' of harmful content and 'friend'/network expansion prompts that recommend children's profiles to adults – regardless of the risks).

Safety by design is a strategic priority for Government,[6] however Ofcom appears to be ignoring this by focusing heavily on content and takedown, but not requiring companies to implement upstream safety measures.

2. 'SAFE HARBOUR' AND SAFETY MEASURES IN THE CODES OF PRACTICE

The Act establishes that services that follow measures set out in Ofcom's codes will be treated as complying with the relevant duties under the Act (known as 'Safe Harbour').[7] Separately, it also sets out that Ofcom's measures must be "sufficiently clear, and at a sufficiently detailed level… that providers understand what those measures entail in practice."[8]

The draft Children's Safety Code recommends 40 measures, which fail to address all risks set out in Ofcom's 'Risk Register.'[9] Ofcom's approach means where there is insufficient evidence or it is disproportionate to include a measure to mitigate a risk, that risk can remain active until a measure is identified – even if that risk is confirmed in a service's risk assessment.

Further, in its statement for the final Illegal Harms Code, Ofcom states it could not require providers to mitigate all risks identified in the risk assessment as the regulator can "only make recommendations [it is] satisfied it is proportionate, having impact assessed them."[10] This is inconsistent with measures in the draft Children's Safety Code, which do not set precise requirements[11] or are drawn from limited evidence.[12]

What this means in practice:

While the Act does set out that the codes can serve as a 'Safe Harbour' provision, it does not stipulate that safety measures must be prescriptive, that the economic burden of measures take precedence above all other considerations (such as the impact on victims), or that these must be grounded in a high threshold of evidence (which only tech companies themselves have access to):

- **"Micro" measures:** Ofcom has interpreted the Act narrowly, requiring very specific recommendations to which costs to services have to be applied before they can be

[6] Draft Statement of Strategic Priorities for online safety, Priority 1
[7] s.49, Online Safety Act 2023
[8] Schedule 4, Online Safety Act 2023
[9] The 'Risk Register' sets out evidence for how harms manifest relating to the risks set out in the Act. This includes features such as: livestreaming functionalities that allow children to view suicide and self-harm content; ephemeral content that exposes children to violent and sexual content; and recommender systems that send children down 'rabbit holes' of harmful content.
[10] Our approach to developing Codes measures, p. 10
[11] See: Draft Protection of Children Code, pp. 18-19, (Measure GA5)
[12] Vol. 5, draft Protection of Children Code, p. 26, 14.34

included in the codes. The Act does not define measures in this way; Section 236[13] states "any reference to a measure includes a reference to any step or action taken […] to comply with duties or requirements under the Act."

- **Proportionality:** Ofcom's approach to proportionality is primarily economic to avoid imposing costs on companies. Whilst the Act requires services take a proportionate approach to duties, and that services' size and capacity is relevant, the Act also specifies that risk and the severity of harms are relevant. **Ofcom's approach does not balance costs to services against the harms for users, the infringement of their rights and the impact on society** – for example, on the criminal justice system or delivering support for victims. This approach means some measures already used in the industry, such as proactive moderation, have not been included in the codes.

- **Technical feasibility**: Ofcom's measures do not reflect the Act's risk-based approach or its outcomes-based duties, and instead require services to follow specific processes or actions. Ofcom has interpreted this to mean only measures that have been tried and tested by industry can be used, **which will not achieve the Act's fundamental purpose of driving safer practice or encourage innovation in safety features – running contrary to the Government's priorities.**[14] This also undermines the Act's intent to be future-proof and widely adopted. In its latest Illegal Harms Code, Ofcom has watered down requirements for content moderation to only take down illegal content where "technically feasible."[15]

3. AGE-APPROPRIATE EXPERIENCES

Age-appropriate design and experiences recognise the different ways that children develop and their different needs at various stages of development, and embeds this in service design. For example, an older teen may need to access information around driver safety, which may include details, images or videos of car crashes which could be disturbing or distressing for a younger child. These provisions already exist in the offline world – for example the British Board of Film Classification[16] sets ratings for films based on the appropriateness of themes and topics for children in different age groups.

There is a clear expectation in the Act that the risk of harm to children in different age groups must be identified and addressed separately.[17]

While Ofcom is clear that services should include children in different age groups in their risk assessment, it proposes no measures to assign children age-appropriate experiences. Instead, Ofcom has chosen to only protect children from 18+ content harm.[18] **This means the experiences, needs and protections of an 17-year-old and a 7-year-old will be considered the same**.

[13] s.236, Online Safety Act 2023

[14] Draft Statement of Strategic Priorities for online safety, Priority 5

[15] See: Sky News, *'Loophole' in law on messaging apps leaves children vulnerable to sexual abuse, says NSPCC*

[16] British Board of Film Classification, About classification

[17] s.12, Online Safety Act 2023 and Schedule 4, Online Safety Act 2023

[18] See: Vol. 5, draft Protection of Children Code, p. 102, 15.319

In certain circumstances, Ofcom relies on the use of age assurance (age verification and/or age estimation technologies) to prevent children from accessing the most harmful content. However, Ofcom has chosen not to apply this to children in different age groups owing to a lack of "independent evidence"[19] on age assurance applied in this manner.

What this means in practice:

Ofcom has failed to meet the purposes or text of the Act. The argument that there is "limited evidence" is without basis – research into how tech companies have complied with the Information Commissioner's Age Appropriate Design Code points to a number of changes that services have made to make their offering age-appropriate.[20] As such, Ofcom's approach already falls below existing best practice, allowing tech companies to 'roll back' on age-appropriate design.

Fundamentally, **Ofcom's approach overlooks the need for age-appropriate safeguards that are tailored to different developmental stages**.

4. LOOKING AHEAD

Ofcom has stated it intends to build on the first iteration of its codes "as [they] gather more information on how the risks of harm to children online evolve and how [its] proposals are impacting this."[21] However, much of this evidence already exists following several years of pre-legislative scrutiny, parliamentary debate, calls for evidence and consultations. As such, it is inaccurate of Ofcom to suggest that there is not a solid foundation of existing evidence.

Given the scale of work required to produce its codes and guidance, it is unlikely the second iterations of these codes will be finalised for some time. Whilst Ofcom will begin consulting on new measures for its Illegal Harms Code in the Spring, **Ofcom must ensure the next iteration of its Illegal Harms Code is substantially more robust** and **the first version of its Children's Safety Code reflects parliamentary intent** to ensure children are protected in the online world.

Considering the Secretary of State's limited powers to require revisions and in the absence of a dedicated parliamentary committee on digital regulation, the crucial early years of this world-leading regulation risk being defined as lacklustre unless changed – having failed to stretch online services as intended.

For more information, please contact reece@5rightsfoundation.com.

Organisations supporting this briefing

5Rights Foundation, Center for Countering Digital Hate (CCDH), NSPCC, Online Safety Act Network, Clean Up the Internet, Internet Watch Foundation, Molly Rose Foundation, Internet Matters, Lucy Faithfull Foundation, Thomas William Parfett Foundation and Global Action Plan.

[19] Ibid., pp. 101-102, 15.317

[20] Children and Screens (2024) *UK Age Appropriate Design Code: Impact Assessment*. See also: Wood, S. (2024) *Impact of regulation on children's digital lives*, Digital Futures for Children Center, 5Rights Foundation, London School of Economics and Political Science (LSE)

[21] Vol. 5, draft Protection of Children Code, p. 4