Stakeholder roundtable

Summary of feedback from civil society organisations and academia on draft industry standards for Relevant Electronic Services and Designated Internet Services

March 2024

This deidentified summary is based on a compilation of feedback received during a roundtable held with civil society organisations and academia stakeholders on 8 December 2023, under Chatham House Rules[1].

Detecting known child sexual abuse material and pro-terror material

- A stakeholder discussed their concerns about the application of the technical feasibility exception in the draft standards. The proposed 'detect and remove' provisions would require providers to deploy tools to automatically detect and remove known child sexual abuse material and pro-terror material where technically feasible. The stakeholder expressed concern that a **technical feasibility test is a higher bar than a reasonableness test** when determining whether a provider can implement mechanisms. They suggested this provision be amended to have **technical feasibility considered as part of a test of what is reasonable and proportionate** for services to implement.
- A stakeholder expressed concern that there is **no explicit statement in the draft standards that the standards will not undermine end-to-end encryption or build weaknesses or vulnerabilities into end-to-end encrypted services**. They stated that while eSafety's public communication has clearly conveyed this, there **need to be protections for digital privacy and security in the standards**. Their concerns included that in the absence of an explicit statement, the standards may be interpreted to incentivise platforms to develop services that will not use end-to-end encryption, or that encrypted services would be required to use tools which were in direct conflict with a service being end-to-end encrypted.
- A stakeholder queried **what body or organisation will make the determination on what is technically feasible**. They highlighted that from their experience, **technology companies may assert something is not technically feasible to provide a barrier or limitation between them and the legal obligation**. Accordingly, they queried whether the company itself or an independent body of experts will determine technical feasibility.
- A stakeholder proposed **greater clarification in the standards around distinguishing platforms that have a greater number of child end-users compared to adult end-users** to ease the concerns around privacy and end-to-end encrypted services. The stakeholder submitted that a **platform which has millions of children and adult end-users who can co-mingle should not be an encrypted environment**.

[1] Participants are free to use the information received, but neither the identity nor the affiliation of the speaker(s), nor that of any other participant, may be revealed.

Disrupt and deter known and new child sexual abuse material and pro-terror material

- A stakeholder suggested the draft standards require **greater clarification of eSafety's expectations under the obligation to disrupt and deter child sexual abuse material and pro-terror material**, and that as currently drafted, it created a **potential risk of over or under enforcement**. They expressed the view that **encouraging platforms to adopt new and unverified tools could be problematic and cause unintended harm**. They proposed a **disclaimer clause** in the standards acknowledging that there are risks with adopting new technologies. Further guidance on what tools can be implemented was also requested.
- A stakeholder expressed the view that **machine learning is unreliable for detecting consent, a user's age or child exploitation**, as human and racial bias can limit the effectiveness of these systems, particularly when identifying images of people who look under 18 years old. They suggested such technology relies on stereotypes of highly variable body characteristics, and machine learning models also cannot scan for consent or exploitation as this is highly contextual. The stakeholder suggested that **rather than having 'disrupt and deter' obligations, there should be greater emphasis on empowering users** to tailor their own moderation.
- Stakeholders suggested **clearer distinctions in the draft standards between the 'detect and remove' obligations and the 'disrupt and deter' obligation**. Further, they propose that the 'disrupt and deter' provision could be **amended to focus on a more holistic approach which educates users** rather than relying solely on content removal. For example, platforms can take proactive steps to cultivate consent culture.

Application of the Relevant Electronic Services Standard

- Stakeholders recommended the **definition of a child in the draft standard have greater cohesion with the Online Safety Act and the United Nations Convention on the Rights of the Child**. The stakeholders expressed concern about the two definitions of 'child' and 'young child' in the draft Relevant Electronic Services Standard. They said having **obligations that only apply to a 'young child' (someone under the age of 16 years old) creates inconsistencies with pre-existing reporting requirements** on child abuse material that is inclusive of children up to 18 years old.
- Similarly, stakeholders noted that **the obligation to set accounts to private by default under the draft Relevant Electronic Services Standard should apply to any child, not just a 'young child'**. This would align with their proposed suggestion to have a blanket

definition in the standards of a child being anyone under the age of 18 and provide safeguards for users up to 18 years of age. This would be consistent with the UK's Age-Appropriate Design Code.

- A stakeholder expressed concern that the draft standards **do not provide adequate protection for children who are not Australian end-users but are being abused online by Australian end-users**. They noted the technical feasibility provision outlines whether it is reasonable to expect providers to incur the cost of taking a particular action, having regard to the online safety risk to end-users in Australia if that action is not taken. **They propose broadening the assessment to ensure the online safety of other children is safeguarded where there is an Australian link.**

Application of the Designated Internet Services Standard

- Stakeholders discussed the **categorisation of pornography sites as a tier 1 high risk category in the draft Designated Internet Service Standard**. They queried why eSafety considers **pornography sites to be higher risk for hosting child sexual abuse material than other services**, and suggested there is often an **inflated figure of child sexual abuse material purported to be on these sites due to stigma**. It was also queried why pornography sites are being included in the draft standards and not just in phase 2 of the codes addressing class 2 material.
- A stakeholder highlighted the need for **greater clarity regarding the application of generative artificial intelligence (AI) categories in the draft Designated Internet Services Standard, and which requirements a service provider will comply with if they offer multiple services at different levels of the supply chain**. They noted some services could be categorised in several of the generative AI categories as they can be an upstream developer, while deploying a model direct to consumers and providing open-source models.
- A stakeholder expressed the view that **sites producing synthetic pornographic material or avatars often use material pirated from dating or porn website databases and the Designated Internet Services Standard could have more consideration for protecting labour and copyright**. Noting that content moderation is difficult due to the bias machine learning tools can have, a stakeholder suggested focusing on prevention around how images used to produce deepfakes have been obtained.

Compliance

- A stakeholder raised whether there should be a requirement for a provider to receive **a direction to comply under the draft standards before enforcement action is taken to align with the industry codes**. They expressed concern there is **potential for abuses of power** to occur if the standards do not provide for due process, and for eSafety to give companies a direction to comply with a requirement prior to the Commissioner using enforcement powers under the Online Safety Act.

Other issues

- A stakeholder proposed the **'policies and terms of use to be published' provision in the draft standards should include the requirement for 'child friendly policies' in addition to plain English versions**, as currently proposed to be required in the standards. They queried **whether child-friendly complaint mechanisms have been considered** in the provisions of the standards that refer to end-user reporting mechanisms.