Consolidated Industry Codes of Practice for the Online Industry

(Class 1A and Class 1B Material)
Briefing session - second round of consultation
21 March 2023, 12:00pm - 1:00pm AEDT

Summary of Discussion

Acknowledgement of Country
Formalities (welcome, introductions, housekeeping)
Introduction of topic / overview the second round of consultation
Recap of key changes to the revised Codes (also refer to Explanatory Memorandum availabe at onlinesafety.org.au)

NB: The session progressed into an open Q&A format. A formal list of questions was not issued in advance as the nature of this briefing was to facilitate an open forum for any questions related to revisions of draft Codes and support participation in the second round of consultation.

Questions and issues raised in discussion, noting Chatham House Rules

Categorisation of RES services in revised draft Codes

- Question raised as to whether the different categories in the revised RES will hold different content removal obligations under the Codes.
    - The categories are more closely related to technical and legal capability, from which different removal obligations arise.
    - The Codes built on the concept included in the OSA which notes not everyone will be able to make context based judgements and, therefore, be able to remove content.
    - Categories for services that are relevant under the *Online Safety Act 2021* (OSA) are included (i.e., have their own category), as well leaving provisions for future services that might need to be assessed (i.e., risk tiering).
    -

End-user classification

- Question raised as to who would be classified as an end-user in the case of an uploading userand 'receiving' end-user (e.g., in a sole trader-client relationship) - would both be end-users or would only the customer?
    - Anyone who can access the content classified as an end-user.

Proactive detection

- Question of which Codes include proactive detection measures.
    - RES, DIS, SMS. Also noted that there are a variety of methods to conduct proactive detection i.e keyword searches, pre-moderated content, image-based automated detection.
- Question of how proactive image detection would be conducted and if it would relate to a database of known content. Also, whether removed content would be added to existing and potentially new databases.
    - Images would be matched to existing databases and could also be added to new databases. The kinds of content that will be included in proactive detection include known

CSAM, new CSAM material (also known as 'first generation' CSAM), known pro-terror material, new pro-terror material (also known as 'first generation' pro-terror material).
    - The most reliable source for known CSAM is the NCMEC database, which can be used with a great degree of reliability.
    - DIS does not include a requirement in relation to proactive detection for first generation material.
- Concerns raised that greater risks are posed for adult content creators who are gender non-conforming individuals or trans individuals. These individuals are in a marginalised demographic and there are concerns their content could be erroneously identified as CSAM, they could lose access to platforms/websites, and be cut off from their income source. Raised that this has been seen under similar schemes internationally.
    - Noted that these measures are intended to target CSAM specifically and not broader categories of adult content.

Appeals processes

- Concerns raised around how nuanced the approach to content removal will be at implementation and the scope for potential overreach, including erroneous or vexatious content reporting, unsophisticated automation models, or ideological differences in interpretation of harmful content.
- Questions around consideration of human based reviews and appeals processes. Concerns related to automated reviews without human context assessments particularly emphasised. Noted that, in financial services under the *Anti-Money Laundering and Counter-Terrorism Financing Act 2006* (AML/CTF Act), there is human review before any action is taken related to a discovered risk.
    - This is included in the guidance on deployment of the technology and there is strong agreement that any deployment of proactive detection needs to keep this in mind, especially considering that the classification system requires context based judgement.
- Also noted particular risks for those whose livelihood is dependent on adult content. Question of what recourse would be available for an adult content creator that is erroneously identified as CSAM and at what point this content would be added to existing databases and hashed. Parallels drawn to robodebt and the impact of automated and incorrect assessments on vulnerable groups.
    - The draft Codes do include some carve-outs related to pornography; however, it is difficult to ensure Codes that are both watertight on risks related to potentially harmful content and sufficiently flexible. The drafts attempt to strike an appropriate balance and organisations or individuals with differing views are encouraged to make a submission to that effect.
    - The nature of the Act errs towards removal if sufficient risk is posed.
    - Further operational experience is needed to fully understand these individual use-cases.

Review of Codes

- In the context of the discussion around concerns in relation to appeals processes and redress, concern raised about whether two years is an appropriate timeline for review of the Codes, with note that the potential scope for false detection and takedowns is currently unknown.
    - Associations responsible for Code development see the risk for the sex industry as limited given that the use of proactive detection via automated means is not mandated in DIS.
    - If systemic issues emerge prior to the two-year mark, there could be a question of whether an earlier review is warranted. All Codes have a mechanism for complaints against the Code itself.

- Raised that an earlier review of the Codes could potentially lend further community acceptance and legitimacy as it would help demonstrate that the Code is being appropriately implemented and effective.
- Question raised of who these complaints would be made to - eSafety, platforms, another body?
    - Complaints would be made to the service provider who would then report these to the eSafety Commissioner. [Ex-post note from authors: complaints would be made to eSafety for DIS.]