Consolidated Industry Codes of Practice for the Online Industry

(Class 1A and Class 1B Material)

Expert Stakeholder Roundtable

13 September 2022, 10.00am- 12.00pm

Summary of Discussion

Acknowledgement of Country

Formalities (welcome, introductions, housekeeping)

Introduction of topic / overview of the Codes with regards to the relevant issues

For reference, the below questions were circulated among participants prior to the Roundtable to guide discussion.

Discussion questions

Structure and scope

1. *What are your general views on the structure and scope of the draft Codes?*

Proactive detection

2. *What should be the scope of services that are required to implement measures to proactively detect material under the Codes? Should there be a different approach between more public services (i.e. services where users choose to make their communications visible to a wider audience) compared to more private services?*
3. *What should be the scope of content that is subject to those proactive detection measures?*
4. *What are the risks and benefits in requiring proactive detection by online service providers? What safeguards, if any, do you think are required where this technology is deployed? If safeguards are required, how should these be included in the Codes (or elsewhere)?*
5. *If proactive detection technology is required to be adopted, are some types of solutions or technology preferred over others?*

Classification and approach to risk

6. *Referring to the guidance provided in the Heads of Terms document about the types of content that falls within the Class 1A and 1B categories, what are your views about the draft Codes' approach to the scope of content that will be regulated?*
7. *Should any additional changes be made to the scope of the Codes or guidance provided about Class 1A and Class 1B materials ?*
8. *The Codes propose a range of different measures for different categories of participants. Some categories of services are treated as having an equivalent risk profile such as search engines, and will be required to introduce the same kinds of measures. Other categories are treated as having different risk levels based on their functionality, scale and purpose; participants in those categories will need to assess their risk level to determine which measures will apply to them. What are your views on this approach?*

Noted from general discussion noting Chatham House Rules

Roundtable participants made the following discussion contributions:

Scope / legitimacy:

- Legitimacy of scheme depends on how all participants can distinguish between socially unacceptable material and false positives. Noting machine learning/algorithms have a degree of inaccuracy. Hence appeals mechanisms are important, as well as ability to hold industry to account through non-government institutions, academia and research
- Codes are missing reference as to how access to data for researchers will be facilitated.
- Noting concerns of scope: inclusion of Class 1B (and later Class 1C) creates issues of legitimacy. The solution does not lie in the creation of many different carve-outs but in an approach that separates obligations for legal material from those for illegal material.
- Broad scope was a mistake from the outset.
- Lack of transparency around complaints (and redress) and how those are addressed by platforms. Inability of platforms to appropriately and timely deal with a false positives (once identified as such) rather than problem of machine learning. This undermines public confidence.
- Conflation of CSEM material and legitimate adult content – arcane system classification of bodies in Australia, which can be misconstrued by machine learning and is prone to misclassification. The proposed scheme makes it difficult for adult industry to operate and advertise. Noted that BDSM material can fall under Class 1A/B material due to inclusion of crime and violence in definition of these classes. This is a problem for the adult industry.

Structure:

- Codes difficult to understand from a consumer perspective.
- Development of Tiers:
    - Noted that tiers were developed on the basis of the assumption that different risk profiles exist (see Position Paper by eSafety). However, given variety of online sections and very broad definition of sections, i.e. broad scope of services within respective sections, there are in reality a range of very different risk profiles. (See further below.)
    - Query: example: a service meets all criteria for Tier 3 but has an extremely high number of users in Australia. How does that impact the risk assessment? Remains a judgement for the organisation. It was noted that assessments must be documented and can be requested by eSafety.
- Concern that even with Tier system, Codes favour larger, established and commercial platforms, to the detriment of community-led and hosted spaces which will have an unreasonable compliance burden. Concern that smaller, independent companies or community groups will be forced out of market thereby worsening competition issues and further consolidating market power/share of major players.

Proactive detection:

- Concern raised that there are some services that are currently used for distribution of CSAM would fall under Tier 2 in social media service Code. Invitation to share details and evidence with the Code authors.
- Concern over limited and costly options for procuring auto-detection tools. Difficulty of understanding how well they work. Concerns that Codes will reinforce the position of dominant players rather than encourage new entrants due to the costs for tools. Concerns shared due to costs to comply with fears that smaller community-led

organisations would exit the market, concentration of market powers, leading to a diminished diversity, freedom of expression etc.

- Others felt that, once regulated, the market would respond and produce cheaper auto-detection options. They also noted that companies currently do not spend enough on human resources to detect and remove material, thereby driving need for improved technology.
- Noted that auto-detection tools work well in limited categories but are also error-prone. They trigger over-posting of material when a party is seeking to draw attention to their content, i.e. causing an extension of more extremist and violent content.
- Difficulty of studying hash data bases as they are secretive. This creates potential concerns if they are required to be an element of online regulation. Others noted concerns that little was known how many databases there were, where they are held, or how they are being used. A much larger global governance challenge.
- Question as to how Codes will disincentivise wrongful removal of adult content? – Codes asks companies to consider a range of factors including human rights impact.
- Criticism that age assurance measures, e.g. DOB collection, are insufficient as a basis for some measures. It was noted that industry is awaiting the outcome of the age verification roadmap, also second set of Codes on Class 2 material.
- Codes are lacking detail around consequences of incorrect action, i.e. false positives as well as failing to detect/remove material. In the former case, what are the avenues for redress or even compensation? Lengthy and costly court processes are not a feasible solution.
- How do we balance missing out on detecting some material in return for not having too many false positives and over-removing? Cautioning against the primacy of wanting to detect all material (noting the broad definition of such material).
- Noted that one of the reasons that society is comfortable with false positives on CSAM is that certain adult content, accounts and livelihood will be a big proportion of the collateral damage, and society is very comfortable with pushing that part of the online world offline.
- Concern of lacking accountability in Codes with respect to false positives. Noted that mandatory reporting would reveal a huge number of false positives. Others noted that false positives are reaching the limits of what consumers are comfortable with.
- In other areas, platforms have made significant commitments with respect to accountability which ought to be translated into best practice requirements for the Codes. Others again reiterated concerns that this would be overly burdensome for smaller players and the chilling effect on competition, diversity of the market, free speech etc. It was noted that Government support would be required if such requirements were placed on such smaller players.

Benchmarking against international approaches:

- Excellent work by NCMEC, similar work in Canada. The recent European initiatives differ from the Australia's expectations in that:
    - Individual providers (no blanket coverage of all industry participants) suggest measures they will take to detect CSAM material to their respective national authority. If the national authority finds that a significant risk remains, a court or independent administrative authority can make a detection order.
    - Detection orders are for a limited period of time;
    - Detection orders are subject to strict procedural safeguards and target a specific type of offence on a specific service;

  - Companies having received a detection order will only be able to detect content using indicators to identify CSAM, provided by the EU Centre;
  - Detection technologies must only be used for the purpose of detecting CSAM, i.e. not any other material;
  - Providers will have to deploy technologies that are the least privacy-intrusive in accordance with the state of the art in the industry, and that limit the error rate of false positives to the maximum extent possible; and
  - The EU Centre to prevent and combat child sexual abuse will maintain a database of indicators for the reliable identification of CSAM.
- Queried operational aspects for global companies to comply with different regulatory regimes.

Classification and approach to risk:

- Question around scope of material and risk of different categories of services. Some expressed that they were happy with the scope and felt it was quite well done how different industry members had become involved. Others noted a focus on Tier 1 services, with the remainder of services potentially skewed to Tier 3 services. Noted the importance of correct Tier assessment.
- Question as to why not same minimum measures across entire eco-system? Explained that differences in the eight online sections are so substantial that this approach proved infeasible. (For example: ISPs all provide uniform service with same risk, hence indeed all have the same minimum measure. However, relevant electronic services (RES) including over-the-top (OTT) messaging services, gaming, texting (via phone), voice calls, etc. provider very different services but yet are categorised within the same online section and have very different risk profiles and abilities (technical and legal) to influence the content, e.g compare OTT messaging with a telco provider's ability to influence a voice call or text message. Similar again for equipment, etc.) In summary, differences within online sections and across online sections necessitate differentiated Tier and risk approach with resultant different minimum measures for Tiers and across sections.
- Noted risk of inconsistent approach to measures, i.e. some companies may decide to assess themselves more conservatively than others, and some may decide to take more action with respect to implementing measures than others. It was noted that the Codes attempt to roll out a greater and improved degree of action at a large scale, in terms of content and coverage of industry participants. (Taking an offline scheme into an online world.)
- How does Head Term impact on TOR or encryption. The Head Terms include limitations that say that services are not required to undermine encryption or other security mechanisms. How does that work in practice? – TOR and use of VPNs etc are not subject to specific measures.

Consultation / process:

- Insufficient time to respond to consultation acknowledging that industry's pressure to deliver Codes in time to Commissioner does not allow for meaningful longer timeframe, especially with young people and NFP organisations. Some noted an unnecessary rush overall in the process (including the development of the Act).
- Upon receipt of all submissions, consideration of submissions and development of final draft for submission. Codes are due for submission for registration on 18 November 2022.
- It was noted that it was expected that the eSafety Commissioner will make a decision on whether to register the Codes prior to Christmas.
- Second set of Codes on Class 2 material to be commenced in the first half of 2023 (?).

Measure of success:

- Periodic review of the Codes. (Initially two years, then every three years.) Question what these reviews will focus on most strongly? E.g. which measures of progress or impact will be examined in depth every couple of years?
- Commissioner decides whether the Codes will be registered. Online Safety Act provides criteria that need to be fulfilled for registration, importantly whether the Codes provide sufficient 'community safeguards'. If the Codes are found deficient, upon initial assessment prior to regulation, the Commissioner can make Standard following specific processes. It was noted that the Commissioner can also make a Standard if the Codes were found deficient once registered.
- Commissioner can register (or decline to register) Codes individually.
- It was noted that it was common regulatory practice for the regulator to determine whether a Code met the objectives and whether it was enforceable. Once registered (if determined that it met the objectives, including appropriate community safeguards), the regulator's role moved to compliance and enforcement. It was noted that this was how the ACMA and ACCC worked. It was confirmed that this was the understanding of the regulatory approach also for the Codes.

Overlap with other legislation / processes:

- Query how the Codes would deal with an amendment of the Classification Scheme. Noted that it was believed that this would not be a short-term process. The 2-year review of the Codes could incorporate changes, if any had been made. Same goes for changes to the Privacy Act and Model Defamation Provisions.
- Noted overlap with requirements of the Telecommunications Act and Telecommunications (Interception and Access) Act which prohibit disclosure and/or detection/interception of communications which is particularly relevant for RES.

Implementation and enforcement:

- If registered, the Codes provide for a 6 months implementation period. In addition, as currently drafted, companies are given an additional 6 months grace period if they can demonstrate genuine efforts to comply with the Codes but have not quite managed to comply with all measures. Noted that this is a period to cooperate with eSafety on compliance. Others voiced scepticism regarding eSafety's willingness to cooperate with industry.
- Noted desire for potential update on regulatory guidance.