Submission to the steering group of industry associations

on the draft Consolidated Industry Codes of Practice for the Online Industry, Phase 1 (Class 1A and Class 1B material)

7 October 2022

Digital Rights Watch is a charity organisation founded in 2016 whose mission is to ensure that people in Australia are equipped, empowered and enabled to uphold their digital rights. We stand for Privacy, Democracy, Fairness & Freedom in a digital age. We believe that digital rights are human rights which see their expression online. We educate, campaign, and advocate for a digital environment where individuals have the power to maintain their human rights.[1]

[1] Learn more about our work on our website: https://digitalrightswatch.org.au/

General remarks

Digital Rights Watch (DRW) welcomes the opportunity to provide a submission in response to the Phase 1 Draft Consolidated Industry Codes of Practice for the Online Industry (the Codes). We recognise there are genuine challenges regarding the safety of vulnerable groups, including children, as well as the distribution of unlawful material online. We also recognise the legitimate interest of the Australian government to promote safer online services to individuals across Australia.

As a leading Australian organisation working to protect our collective digital rights, DRW is primarily concerned with ensuring an appropriate balance is struck with regard to the impact upon individuals' and communities' rights, including any adverse impacts it may have on privacy, digital security, and freedom of speech and expression.

The key areas of concern we raise in this submission are:

1. the Codes interaction with the government's ongoing reform agenda, including pending reform to the Privacy Act and review of the National Classification Scheme,
2. the risks and challenges of proactive detection of material, and that its use should be carefully and strictly limited and with robust safeguards to prevent over- or mis- use,
3. the lack of clarity regarding coverage thresholds, and possible adverse impacts upon competition which ultimately consolidates power into the hands of large commercial entities, and
4. the need to include provisions to increase accountability of both industry as well as the eSafety Commissioner by way of providing access to reporting and data for public interest and research purposes.

Digital Rights Watch has played an active role in previous consultations regarding the Online Safety Act which remain relevant to the draft industry codes, including submissions to:

- the proposed *Online Safety Bill 2020*[2]
- the draft Basic Online Safety Expectations[3]
- the draft Restricted Access Systems (RAS) Declaration[4]

Other relevant submissions DRW has made in the past 18 months include to the:

- Parliamentary Inquiry into Social Media and Online Safety[5], and the
- Parliamentary Joint Committee inquiry into the *Criminal Code Amendment (Sharing of Abhorrent Violent Material) Act 2019.*[6]

---

[2] Digital Rights Watch submission, February 2021. Available here: https://digitalrightswatch.org.au/2021/02/18/submission-the-online-safety-bill/
[3] Digital Rights Watch and Global Partners Digital joint submission, November 2021. Available here: https://digitalrightswatch.org.au/2021/11/04/submission-draft-basic-online-safety-expectations/
[4] Digital Rights Watch submission, November 2021. Available here: https://digitalrightswatch.org.au/2021/11/25/submission-draft-restricted-access-systems-declaration/
[5] Digital Rights Watch submission, 12 January 2022. Available here: https://digitalrightswatch.org.au/2022/01/13/submission-inquiry-into-social-media-and-online-safety/
[6] Digital Rights Watch submission, October 2021. Available here: https://digitalrightswatch.org.au/2021/10/29/submission-abhorrent-violent-material-act/

1. Pause further development until the Codes can be aligned with the government's ongoing reform agenda

We have previously raised concerns regarding both the timing and the consultation processes throughout the development of the entire online safety regime to date. Unfortunately, the Codes follow the alarming trend of only providing civil society and other relevant stakeholders with an extremely short timeframe to provide feedback on a complex regulatory scheme. These proposals stand to have significant impact upon the way that people are able to use digital services as well as their fundamental human rights, and require appropriate time for community groups, small businesses and civil society organisations to consider and meaningfully respond.

In addition to this, the Codes are being developed ahead of significant regulatory reform that is highly likely to impact the way the Codes are implemented. While we understand the desire to move quickly, doing so runs the risk of creating an overly complex, contradictory, and constantly changing regulatory landscape.

The Codes should be consistent with broader government policy and related regulation. For instance, the outcome of the review of the Privacy Act is likely to have a direct impact upon the Codes. Personal privacy is crucial to online safety for many people, especially vulnerable populations, and so any industry code providing guidance for online safety must integrate best practices for privacy protection.

Similarly, the Codes, and indeed the entire approach of the Online Safety Act, is based upon the controversial and outdated National Classification Scheme, which is also currently under review. Developing an approach to regulation that is based upon an outdated and soon-to-change foundation is a surefire way to exacerbate Australia's already fragmented and complex internet and communications regulatory regime. We suggest that further development of the Codes is paused until such a time that the Codes can align with, and give effect to, relevant legislation currently under review, including the Privacy Act and the National Classification Scheme.

2. Proactive detection should be carefully and strictly limited

We were pleased to read in the explanatory memorandum that the industry associations "concluded that the extension of proactive detection measures could have a negative impact on the privacy and security of end-users of private communications and file storage services, including services used by businesses and government enterprises." We agree that proactive detection of material in private communications and file storage is an unreasonable invasion of privacy and creates additional security and safety risk for individuals, businesses and governments.

Further, we appreciate that industry associations have also given consideration regarding the detection of first-generation child sexual abuse material (CSAM) and determined that detection should be limited to that of *known* CSAM only. While we understand the need to address harms associated with the distribution of first-generation CSAM, the technology to detect material that has not been previously identified currently presents unreasonable challenges and risks with relation to accuracy, over-capture, privacy and security.

Proactive detection will always carry with it some level of privacy and security risk. While Digital Rights Watch does not argue against the use of hash scanning in public platforms for known CSAM, we remain concerned that there are not adequate safeguards in place to prevent use of proactive detection technology from expanding into other areas. Given the inherent risks to privacy and digital security, any requirement placed upon companies to use proactive detection technologies must be carefully balanced with robust safeguards and restrictions to prevent misuse and abuse of technology.

We further note that while we appreciate the distinction made such that private file storage and communications are not included, we remain concerned that in practice, differentiating between images shared on a public forum as opposed to within private communications may not be technically feasible.

There is nothing within the enacting legislation, the *Online Safety Act*, that creates obligations on service providers to proactively monitor communications over their networks. This was a key part of the policy debates in the lead-up to the passage of the Act. It is not acceptable for the Codes to be used as a way to extend obligations beyond the legislative intent reflected in the Act—this would represent a serious overreach of eSafety power.

We also note that the EU *Digital Services Act* has introduced a prohibition on general monitoring. This was developed in response to concerns regarding some policymakers asking platforms to scan all communications to find particular content. We wish to highlight that there is a risk of creating challenging misalignment between Australia and international jurisdictions should Australia move toward requiring general monitoring, while the EU prohibits it.

Challenges of automated processes to detect unknown or previously unseen content
High quality automated detection of harmful content is extremely difficult, especially when the scope of target content is broad, as is Class 1B material.

Machine learning classifiers which seek to automatically flag possibly harmful content (as opposed to matching content to known, or previously identified harmful content) may be improving, but they remain seriously flawed when it comes to classifying complex material at scale. One issue is accuracy and the risk of both over- and under- capture of content. We note our support for groups such as Scarlet Alliance and Assembly Four which have emphasised the harm caused to sex workers and others who post legal sexual material when their content is taken down due to incorrect or overly broad content classification.

Another issue is that due to the lack of training data, classifier models are more likely to make mistakes related to marginalised groups, and in doing so further entrench existing inequality. Further, there remain ongoing challenges regarding the explainability of machine or deep learning classifiers. While this is an area of ongoing technical research and development, at this stage it may not be possible to explain or justify why some content is flagged by an automated machine learning content classification system.

DRW recommends that the Codes should not be extended to require monitoring beyond matching of *known* CSAM in the highest risk public and semi public services. In addition to this, strict limitations and safeguards should be in place to prevent this technology from being rolled out more broadly.

3. Coverage thresholds need further clarity and risks to competition should be carefully considered

There remain outstanding questions regarding the thresholds for application of the Codes to different providers. We are concerned that this is not sufficiently clear under the current draft of the Codes. It is extremely challenging to evaluate possible digital rights impacts of the proposed Codes without clarity of which service providers will be regulated under which categories.

In addition to this, we remain concerned about the possible harmful impacts of limiting competition by increasing regulatory burden upon small, independent, community-led or non-commercial entities. Despite the three-tiered risk assessment system, there remains a strong incentive for entities to 'round up' their compliance where there may be any confusion regarding which category or tier they might belong to, incurring what is likely to be an unreasonably regulatory compliance burden.

While we understand the urge to compel a larger number of entities, including small ones, toward the more stringent obligations, we wish to highlight that in doing so it is likely to threaten the vibrance and diversity of the internet—which is made up of far more than just the major players.

The Codes currently suit incumbent powerful companies, especially large social media companies, as more risk and compliance cost for community-led and hosted online spaces means that more traffic will be driven to Big Tech. For instance, why would any community group go to the effort when it is less risky to just set up a Facebook group? **Consolidating power and market share of large commercial companies ultimately undermines our collective online safety.**

To be clear, we are not advocating for small entities to avoid all regulation, but that care needs to be given to ensure that the Codes do not penalise or threaten small, non commercial or community-led entities or online spaces. We note that the Office of the

eSafety Commissioner has made some assurances to use discretion when considering liability of smaller entities. We wish to emphasise that this promise is not enough to lower the risk for many small entities.

4. Increase accountability by including provisions for access to reporting data for public interest research

There are currently no provisions in the draft Codes for mandatory reporting or provision of access to reporting data for public interest or research purposes. Instead, reporting is limited to the eSafety Commissioner.

We recommend including provisions within the Codes to require members of the online industry to provide access to reporting data. This assists in research and public interest auditing, ultimately assisting in the transparency and accountability of the regulated entities, the eSafety Commissioner, and regulatory scheme in general. High level aggregated reporting is not enough.

Recommendations

- Pause further development until such a time that the Codes can align with, and give effect to, relevant legislation currently under review including the Privacy Act and the National Classification Scheme.
- The Codes should not be extended to require monitoring beyond matching of *known* CSAM in the highest risk public and semi public services. In addition to this, strict limitations and safeguards should be in place to prevent this technology from being rolled out more broadly.
- Clarify the coverage thresholds to make it clear which services are captured by which category, in order for more comprehensive consultation and feedback to take place.
- Include provisions to require access to data for public interest research purposes, to improve transparency and accountability.

Contact

Samantha Floreani | Program Lead | Digital RIghts Watch | samantha@digitalrightswatch.org.au