GNI Analysis of Consolidated Industry Codes of Practice for the Online Industry, Part 1 (Class 1A and class 1B material)

I. Introduction

The Global Network Initiative (GNI) is a multistakeholder initiative that brings together 90 prominent academics, civil society organizations, information and communications technology (ICT) companies, and investors from around the world. Members' collaboration is rooted in a shared commitment to the advancement of the GNI Principles on Freedom of Expression and Privacy, which are grounded in international human rights law and the UN Guiding Principles on Business and Human Rights (UNGPs). For over a decade, the GNI Principles and corresponding Implementation Guidelines have guided ICT companies to assess and mitigate risks to freedom of expression and privacy in the face of laws, restrictions, and demands, including in politically sensitive contexts.

GNI welcomes the opportunity to provide feedback on the industry codes recently published for comment pursuant to the Online Safety Act. GNI's global and multistakeholder membership is closely following developments in Australia, in no small part because of the important role that Australia plays as a model for democratic governance in Asia, throughout the Commonwealth, and globally. That reputation and influence creates both opportunity and peril.

II. Background

In 2020, GNI conducted an analysis using human rights principles of existing and proposed governmental efforts to address various forms of online harm related to user-generated content — a practice we refer to broadly as "content regulation." This included evaluation of Australia's 2019 Online Safety Discussion Paper. After extensive consultations with GNI members and outside stakeholders, including governments, in a wide range of jurisdictions, GNI published a policy brief titled "Content Regulation and Human Rights: Analysis and Recommendations," ("Policy Brief") which set out a range of observations and suggestions on how to regulate content in a manner that upholds and strengthens human rights.

That analysis informed our May 2021 submission to the Australian government on the then proposed Online Safety Bill. In that submission, we noted serious concerns including: the overly broad and undifferentiated application of the Bill to companies across the spectrum of services; extensive discretionary powers afforded to the eSafety Commissioner; limited exemptions for content in the public interest; an inflexible emphasis on a 24-hr takedown window; and lack of definitional clarity around thresholds for certain categories of content. Many of these concerns continue to inform GNI's analysis of the resulting industry codes that have now been put forward for feedback.

GNI hopes that this feedback will be useful and remains eager to engage with Australian authorities, industry associations, and civil society to ensure that Australia's approach to online safety is consistent with the country's long-standing commitments to international human rights principles, including through its engagement in the Freedom Online Coalition and the recent Declaration for the Future of the Internet. If carefully balanced and subject to appropriate safeguards, including regarding transparency in implementation, independent scrutiny and oversight, and opportunities for adjustment going forward, GNI is hopeful that Australia's approach can help demonstrate effective *and* rights-protecting content regulation. However, absent these safeguards, there is a real risk that Australia's approach will have negative impacts on human rights online, both in Australia and beyond.

I. Scope of application

GNI has concerns that the industry codes recently published for comment pursuant to the Online Safety Act have ambiguity surrounding how the code will be applied to different types and sizes of companies and organizations with a presence in Australia or otherwise affecting Australian citizens.

When considering the scope of application, the eSafety Commission should consider the principle of necessity, which suggests focusing regulation on particular services in order to minimize its impact on expression. As a general rule, the more distant a particular service is from the end user, the less visibility and granular control it has over user-generated content. Even for services that are "close" to end users, it is important to consider a variety of factors, including the type of service and its functions, the extent to which user generated content is public or private, and the extent to which content or data are persistent or ephemeral. As GNI noted in the Policy Brief, lawmakers and regulators would be well served to carefully consider which types of private services, at which layers in the ecosystem, are most appropriately positioned to address the specific concern(s) at issue and to constrain their approaches to those best positioned to address those concerns.

The type of and proportional impact that the codes will have on non-profits, start-ups and smaller entities should also be taken into consideration. Requirements to regulate speech may have unintended impacts on the pluralism of content and providers of consumer services that may be available. Of particular concern is that the introduction of any OP Code provisions may require ICTs to collect more information than they otherwise would for that entity's functions or activities.

Finally, it is important to understand the ways in which industry codes may affect the Internet ecosystem at large — including research, public archiving, historical, artistic, and journalistic activities. For example, prescriptive notification requirements that, in practice, require

additional instances of notification by ICTs can lead to ICTs being forced to collect information that they would not ordinarily do so as part of their business practices. It may also be unduly burdensome on nonprofits, start-ups and smaller entities. GNI recommends tailoring the Codes based on the impact and capacity of an entity, since the number of users can be a poor proxy for the actual impact a platform has on the underlying content or conduct at issue.

II. Transparency, Accountability, and Consistency

The power of the eSafety Commissioner to unilaterally revise the industry codes raises important questions about accountability and transparency. Industry associations developed codes in keeping with guidelines in the eSafety Commissioner's September 2021 position paper. However, after the codes are submitted, the eSafety Commissioner retains broad powers under the Online Safety Act to impose its own industry standards if the Commissioner determines that the submitted codes are deficient. Where there is a discrepancy between the eSafety Commissioner's position paper and the industry codes—such as on technological tools for proactive detection—it is unclear which view will prevail in the final and binding version of the industry codes. We recommend a system of review and oversight to ensure that the eSafety Commissioner's revisions of industry codes are consistent with principles of human rights and democratic governance.

It is also unclear whether the provisions of the industry codes will soon become obsolete or inconsistent with other Australian legislative projects. Given that the ongoing review of both the Privacy Act and the classification system that undergirds the industry codes may lead to changes in both of those policies, related regulations in the industry codes may soon become redundant or contradictory. For example, the code's guidance on harmful content subject to the code may need significant revision if the classification system is overhauled. Similarly, rules about monitoring user-submitted content and identity verification for certain services may conflict with future privacy regulation. We therefore recommend that a review process ensure regulatory coherence and clarity between the industry codes, privacy legislation, and revised classification system.

III. Proactive Detection of Content

We have concerns about the eSafety position paper's emphasis on the significant role of proactive detection technologies. In our Policy Brief, GNI cautioned against overreliance on automated tools to proactively detect and remove content. Such tools can be flawed and often lack the ability to assess important context, and may lead to unnecessary removal of legal content. This can result in both the under-removal of illegal content, and the unwarranted removal of legal content. In addition, there is a significant risk that the error rates and impacts of such tools will fall disproportionately on marginalized communities and voices, who are also less able or willing to use grievance or appeals mechanisms to correct these mistakes.

Although the industry codes only require certain high-risk services to use hashes and other tools to detect child sex abuse material (CSAM), the codes nevertheless encourage both the development and use of automated tools and processes to detect, report, and remove class 1 material more generally. A range of mechanisms and techniques exist for identifying, verifying, hashing, and sharing CSAM in appropriate ways. However, the same cannot be said for other categories of problematic content. This is in part due to the fact that satirical, humorous, journalistic, and counter-messaging content is much less likely to be confused for CSAM, as has been documented to be the case for violent content or other categories that may eventually be deemed "class 1" material. Given the risks inherent in these tools in these non-CSAM categories, we recommend that the codes acknowledge these distinctions and provide guidance on how to assess and mitigate the risks to freedom of expression associated with proactive detection technologies, including through human review, redress mechanisms, and appropriate transparency.

IV. 24 hour takedown rules

As described in the codes, various services are required to remove content within 24 hours or "as soon as reasonably practicable" when there is "evidence of a serious and immediate threat to the life or physical safety of an Australian adult or child." Although GNI applauds the effort to limit such strict and short timelines to a narrower category of content, the deadline may nevertheless be very tight for services that process enormous volumes of content and could be extremely burdensome for smaller services who lack the resources to monitor and adjudicate content that quickly. Notwithstanding the "reasonably practicable" caveat, there is a significant risk that services will be penalized despite good-faith efforts to evaluate and remove harmful content due to the arbitrary strictness of the timeline. This could lead to an inappropriate reliance on automated detection technologies, notwithstanding their limitations.

GNI's Policy Brief warned that by "imposing strict time limits on all content adjudication, states may effectively hinder the ability of ICT companies to prioritize resources and make nuanced, content and circumstance-specific determinations." We recommend that the narrow approach taken by the codes be preserved and that the eSafety Commissioner and other relevant authorities continue to work closely with all stakeholders to develop shared understandings of the "reasonably practicable" standard for taking down harmful content in emergency situations.

V. Geographic scope

The codes are designed to cover all internet services that are accessible to Australian end-users, which means that the codes could apply to any service, website, or provider in the world,

regardless of their relationship with or physical presence in Australia. This raises a range of jurisdictional questions, including the extent to which the codes will end up impacting the services and content available to users in other jurisdictions, potential conflicts of law that could be created, forum shopping by companies and users, and whether and how companies or services not based in Australia may face consequences, including blocking in Australia.

While section 6.1h of the Head Terms state that the codes do not require breach of foreign laws about managing personal information of foreign end-users, it is unclear whether the potential conflicts between the codes and applicable foreign or international laws have been sufficiently examined, and how contradictions will be adjudicated. We recommend further clarification on how the risks of conflicts between Australian and foreign law should be examined, understood, and avoided or resolved.

VI. Conclusion

As Australia seeks to ensure the safety of its citizens both on and offline through the online safety codes, GNI urges care in ensuring that the industry codes preserve freedom of expression and privacy rights for all users. Given the vast scope of the codes and the wide range of actors in the online ecosystem, regulators should:

- tailor recommendations and requirements for online services based on their purpose;
- limit and clarify circumstances where proactive detection of harmful content is required;
- develop shared understandings of reasonable timelines for taking down content; and
- clarify how contradictions between the codes and foreign and international laws will be adjudicated.

We encourage the industry associations to consider these recommendations as they revise and update the codes for registration. We look forward to further engagement with the industry associations and the Australian government on future industry codes and technology policy development.