Director, Online Safety Research and Reform Section
Department of Infrastructure, Transport, Regional Development and Communications
By email: onlinesafety@communications.gov.au

19 February 2020

Digital Rights Watch submission to Online Safety Legislative Reform consultation

Prepared by Lizzie O'Shea and Nicolas Suzor

About DRW: Digital Rights Watch (DRW) is a nonprofit charity that works to ensure that Australians are equipped, empowered and enabled to uphold their digital rights: https://digitalrightswatch.org.au. For more information about this submission, contact: lizzie@digitalrightswatch.org.au

Preliminary matter: high level objects and the definition of safety

As an initial matter, we believe that the definition of 'online safety' used throughout the discussion paper is too narrow -- and that the objectives of this review cannot be met in isolation without deep engagement with other ongoing reviews and consideration of interrelated issues. We appreciate the effort taken in this review to harmonise some aspects of regulation that applies to online safety, but we note that regulation in this area could benefit from a much more comprehensive understanding of online harm.

Our concern here is that this current review does not sufficiently emphasise the extent to which online safety issues are interconnected with complex issues of cybersecurity and privacy. In particular, the ongoing lack of a private cause of action for serious infringement of privacy, as recommended by many successive reviews, continues to leave Australians exposed to a broad range of serious harms.

We suggest that the involvement of vulnerable users and civil society is critical to making this regime fit for purpose. It should be part of the high level objectives, and appropriate consultation

processes built in to ensure that their unique perspective can be incorporated into any future substantive legislative reform.

We also note the opportunity for this reform to reinforce existing and concurrent policy initiatives such as the 2020 refresh of Australia's Cyber Security Action Plan, which in 2016 had a goal to '*champion an open, free and secure Internet to enable all countries to generate growth and opportunity online*.' This should remain a pillar of any attempt to update key policy within the online world.

Finally, the Discussion Paper notes that voluntary arrangements and liaison with the eSafety Commission have "been successful in the vast majority of cases in relation to Australian-hosted internet content, cyberbullying material and intimate images."[1] In the context in which voluntary compliance is the norm, and working well, we suggest that a substantial degree of caution is warranted before imposing new legal obligations or introducing new legal powers for regulators. To the extent that new powers for regulators and administrative officials are introduced, we strongly urge that they are accompanied by a high standard of transparency and accountability for how they are exercised.

Appropriate thresholds for seriously harmful material

In creating a harmonised takedown regime for online content, it is important that the Government ensures that the definitions of seriously harmful material are limited only to material and behaviour that would be unlawful for individuals to post online. The globally accepted principles of freedom of expression mean that service providers must not be required under law to remove content that is not unlawful.[2]

In developing a takedown obligation, we note that an apparent intention to cause "serious distress or serious harm" with material that is "menacing, harassing or offensive" could catch an undesirably broad category of content. The category of behaviour that is potentially unlawful under s 474.17 of the Criminal Code Act 1995 (Cth) is already unjustifiably broad. The inclusion of "offensive" behaviour creates a much lower threshold for illegality than the common law obscenity offence or other relevant legislated criminal offences.

We support the position in the discussion paper to limit the scope of content to material that is unlawful under Commonwealth criminal law. However, we note that it is difficult to provide comments about the scope of the definition of harmful material at this stage without seeing draft legislative text. This is a difficult area, however, and we would like to stress the importance of careful drafting to avoid overly broad definitions. A scathing negative review, for example, can

[1] Discussion Paper, p 6.
[2] 'Manila Principles on Intermediary Liability: Best Practices Guidelines for Limiting Intermediary Liability for Content to Promote Freedom of Expression and Innovation' (24 March 2015) <https://www.manilaprinciples.org/> ('Manila Principles').

be calculated to cause serious harm (to a person's business interests) and be worded in offensive language, but should not fall within the scope of a takedown obligation unless it amounts to an injurious falsehood. News reports of genocide or torture may rightfully be distressing and offensive, but serve a legitimate public interest, and should therefore not be within the scope of a takedown regime.[3]

We suggest that the definition of cyber abuse is clearly articulated to only target material that is criminal to possess or post. This should be done explicitly through the legislation and supported by explanatory notes for the avoidance of doubt, in order to avoid potential scope creep.

We also strongly suggest that a clear exception be created for material that is in the public interest.

Parliamentary process

We believe that restrictions on speech, whether they are enforced through direct liability or through intermediaries, should only be introduced through full parliamentary processes. This is important in order to ensure that restrictions on speech are compatible with international human rights norms. Accordingly, we do not believe that it is appropriate to delegate the ability to designate new categories of harmful content to the Minister. The quick legislative response to the Christchurch massacre[4] shows that normal parliamentary processes are quick enough to address emerging harms without the necessity of defining harmful material through delegated legislative instruments.

We note that it is also possible to avoid having to use delegated instruments to cover new channels for harm by framing any obligations in a technologically neutral way. For example, the Discussion Paper raises concerns about the future need to deal with 'virtual reality or animated content'[5] -- but these are different *channels*, not new *harms*. We suggest that the ongoing efforts to harmonise media regulation across different distribution channels, including the review of the Classification regime, can provide better insights on how to design legislation that is suitable for a converged media environment without the need to expand the scope of the regime through Ministerial declarations.

---

[3] This is a problem with the current *Criminal Code Amendment (Sharing of Abhorrent Violent Material) Act 2019* (Cth) which, for example, criminalises the hosting of images of torture taken by the perpetrators, like the images of the Abu Ghraib abuse scandal. These are clearly images that are in the public interest, but their distribution is not covered by an appropriate exception to the definition of Abhorrent Violent Material. See, for example, https://en.wikipedia.org/wiki/Abu_Ghraib_torture_and_prisoner_abuse.

[4] *Criminal Code Amendment (Sharing of Abhorrent Violent Material) Act 2019* (Cth)

[5] Discussion paper, p 41.

The importance of review

Any obligation on service providers to remove content must be balanced with an appropriate mechanism to correct mistakes. Generally speaking, when a service provider has an enforceable obligation to remove content, they will often err on the side of caution in cases of uncertainty.[6] It is also clear that even the most carefully administered takedown processes are routinely unintentionally mistaken and subject to deliberate abuse.

A well designed takedown regime should have a clear avenue of appeal for anyone affected by a takedown decision.[7] Any decision by the eSafety Commissioner or any other body to issue a takedown notice or otherwise deem material to be prohibited should be subject to full merits review by a competent court. It may also be desirable to allow an initial appeal to the Administrative Appeals Tribunal as a faster and lower-cost alternative.

Blocking obligations for internet service providers

Blocking content at the infrastructure level is an extremely blunt form of censorship that should only be used in the most extreme cases. It is generally not possible for an Internet Service Provider to block individual pieces of content -- blocking content at an ISP level implies blocking entire domains or IP addresses, which often necessarily blocks access to legitimate content.

We believe that a court order should be required before Internet Service Providers are required to block content. This is the norm established in copyright law, where the Federal Court has reiterated the importance of ensuring that injunctions directed at ISPs are appropriate tailored to avoid an undue burden on freedom of expression.[8]

Where an ISP is required to block content, it is important that an appropriate message is displayed to anyone seeking to access the material, informing them of their rights of appeal.

Jurisdiction and geo-blocking

Australian laws are often different to the laws of other countries. There is an emerging threat that many countries are seeking to apply their laws in an extraterritorial manner. If this trend continues, it will eventually result in the freedoms of Australians routinely being restricted by the laws of other countries in ways that are incompatible with our democratic processes.

---

[6] Kylie M Pappalardo and Nicolas P Suzor, 'The Liability of Australian Online Intermediaries' (2018) 40 *Sydney Law Review* 469.
[7] Nicolas P Suzor, *Lawless: The Secret Rules That Govern Our Digital Lives* (University Press, 2019) ('*Lawless*').
[8] *Copyright Act 1968* (Cth) s 115A.

It is important that any power to require a service provider to remove or derank content be limited to apply only in Australia. Technically, most service providers can use geoblocking techniques to ensure that the effect of Australian law is not felt extraterritorially. Any proposed scheme should explicitly only require content to be removed or blocked for Australian users.

Transparency

Improving transparency practices is fundamentally important to enabling people to evaluate how online services are governed. The Santa Clara Principles on Transparency and Accountability in Content Moderation provide a global minimum standard for acceptable transparency reporting in content moderation.[9] These set out the basic requirements of legitimate content moderation systems, including requirements to publish regular aggregate figures, provide notice to affected individuals, and establish accountable review processes for content moderation decisions. We strongly recommend that government efforts to improve industry transparency reporting practices use these principles as a starting point, in order to promote industry compliance with global best practice norms.

Importantly, aggregate transparency reporting on its own is insufficient to evaluate how well digital platforms are responding to harmful content and behaviour on their networks. Much more information is required in order to better understand how harmful behaviour is perpetrated online, how harmful content is shared and amplified, and how well digital platforms are responding to improve safety. Understanding these issues requires genuine and deep collaboration between industry, regulators, academics, the media, and civil society.[10] We suggest that any effective regulation in this complex area requires careful attention to develop an evaluation scheme that can track the prevalence and nature of threats as well as the performance of different platforms. Specifically, this means that platforms should be encouraged to find ways to provide researchers with much better access to fine-grained data on moderation decisions to enable independent public interest research.[11] Current industry approaches to providing data continue to lag behind what is needed to really tackle online safety issues in a meaningful way.

We urge the Government to encourage more meaningful transparency from industry stakeholders as part of any articulation of basic online safety expectations.

---

[9] https://santaclaraprinciples.org/

[10] Nicolas P Suzor et al, 'What Do We Mean When We Talk About Transparency? Toward Meaningful Transparency in Commercial Content Moderation' (2019) 13 *International Journal of Communication* 1526, https://ijoc.org/index.php/ijoc/article/view/9736.

[11] Axel Bruns, 'Facebook Shuts the Gate after the Horse Has Bolted, and Hurts Real Research in the Process', *Internet Policy Review* (25 April 2018), https://policyreview.info/articles/news/facebook-shuts-gate-after-horse-has-bolted-and-hurts-real-research-process/786.

We also believe that the reporting obligations of the eSafety Commissioner should be enhanced. The eSafety Commissioner should have clear obligations to provide regular and detailed reports about how any new powers are exercised.

Children's Rights in online spaces

Any reform to online safety must have specific and targeted elements that address the nuance of children's rights. It has been welcoming to see the eSafety Commissioner's focus on educating both adults and children on children's safety online, and any legislative reform must continue this focus. The addition of a child-centred approach to this safety education will ensure that individual children's rights are protected and upheld.

Children's rights are enshrined in the Convention on the Rights of the Child (UNCRC), which was adopted unanimously by the United Nations General Assembly in 1989. It has since become the most rapidly and widely ratified human rights treaty in history, and its operationalization is supported by a series of Optional Protocols and General Comments. The UNCRC encompasses a broader range of rights than any other human rights treaty, from humanitarian to economic, and socio-cultural to civil and political rights. While the UNCRC is not the first international treaty to protect children's rights, it stands apart from previous declarations in that it grants children the right to express their opinion in matters that concern them, thus adding participation rights to those of protection and provision that were laid out by the UNCRC's precedents. Australia is a signatory to the UNCRC and is held accountable to the attendant duties and obligations to children by the UN's monitoring and reporting processes.

Children's digital rights have been an explicit concern of the international children's rights community since at least 2014 when, in observance of the 25th anniversary of both the adoption of the UNCRC and the release of the code that would become the internet, the United Nations Committee for the Rights of the Child held a Day of General Discussion (DGD) on 'Digital Media and Children's Rights'. The DGD brought together global experts from across sectors to discuss how to interpret the UNCRC to harness the opportunities and meet the challenges of the digital age. It marked an attempt to seriously consider how to balance children's protection from harm online with promoting the benefits for children of their digital media engagement. Further, the DGD aimed to not only promote children's rights to access the internet safely but also to consider ways digital media might better enable children to understand and enact a broad range of rights in their everyday lives.

A Case for a General Comment on Children and Digital Media[12], commissioned by the Children's Commissioner of England, asserts that, internationally, states, NGOs, and corporates are calling for principled and evidence-based guidance to deliver on children's rights for the digital age. The Australian government should lend support to this General Comment on Children and Digital media, which would guide states, NGOs and corporations in their interpretation of the Convention on the Rights of the Child for the digital age, enabling duty bearers and implementing organisations to prioritise children's rights in relation to digital media.

Any legislative reform that directly impacts children's rights should also actively engage children and young people in developing responses that protect their rights to provision, protection and participation in the digital age, and develop child-centred measures of impact.

Product design

The design of products (other than online products) appears to be referenced throughout the paper but it remains unclear the extent to which the Minister (or eSafety Commissioner) will consider the safety implications of this process. Given the widespread availability of internet-enabled products marketed to children, this ought to be a strong focus of any new legislative regime. We appreciate the interest from the government in promoting an industry that ensures that products marketed to children default to the highest level of privacy and safety, and we think this ought to apply to physical products that are internet-enabled as well as online products like apps. Policies and guidelines could be developed in partnership with the Australian Competition and Consumer Commission to ensure that products promoted to children are fit for purpose and take into account the inherent vulnerability of children who are users. This could include a non-exhaustive list of standard requirements as default, including requirements such as two factor authorisation or end to end encryption (as applicable) with penalties for non-compliance.

Image-based abuse

Image-based abuse is a strongly gendered manifestation of sexual violence.[13] It is important that online service providers continue to work to improve the way they deal with image-based abuse in particular,[14] and gender-based violence in general,[15] perpetrated through their

[12] Livingstone, S, Lansdown, G. & Third, A. 2017. The case for a UNCRC General Comment on Children's Rights and Digital Media: A report prepared for the Office of the Children's Commissioner of England. London, LSE Consulting. Available at: https://www.childrenscommissioner.gov.uk/wp-content/uploads/2017/06/Case-for-general-comment-on-digital-media.pdf.

[13] Clare McGlynn, Erika Rackley (2017) Image-Based Sexual Abuse. *Oxford Journal of Legal Studies*, 37(3), pp. 534–561, https://doi.org/10.1093/ojls/gqw033

[14] Nicolas Suzor, Bryony Seignior, & Jennifer Singleton (2017) Non-consensual porn and the responsibilities of online intermediaries. *Melbourne University Law Review*, *40*(3), pp. 1057-1097, http://law.unimelb.edu.au/__data/assets/pdf_file/0019/2340424/09-Suzor,-Seignior-and-Singleton.pdf.

networks. We note that there are still major problems with inconsistent laws and the prevalence of victim-blaming advice to deal with image-based abuse.[16]

We support the extension of the image-based abuse scheme to deal with images that purport to be a person—the recent rise of 'deepfake' image-based abuse is concerning and harmful.[17]

Governance and participation - the role of the eSafety Commissioner

The functions of the eSafety Commissioner are notably broad, and we support this generally. One aspect of the eSafety Commissioner's work that ought to be considered critically important is involvement of vulnerable users in the process of standard setting and government policy design. This will ensure that such policies will incorporate the perspectives of people who are too easily ignored, especially by industry, until after harm has been experienced. We think the function of facilitating community consultation, particularly with vulnerable groups, is something that the eSafety Commissioner could lead on and provide resources for if it was articulated explicitly as one of his or her functions.

[15] Molly Dragiewicz, Jean Burgess, Ariadna Matamoros Fernandez, Michael Salter, Nicolas Suzor, Delanie Woodlock, & Bridget Harris (2018) Technology facilitated coercive control: Domestic violence and the competing roles of digital media platforms. *Feminist Media Studies*, 18(4), pp. 609-625, https://doi.org/10.1080/14680777.2018.1447341; Nicolas Suzor, Molly Dragiewicz, Bridget Harris, Rosalie Gillett, Jean Burgess, & Tess Van Geelen (2019) Human rights by design: The responsibilities of social media platforms to address gender-based violence online. *Policy and Internet*, 11(1), pp. 54-103, https://doi.org/10.1002/poi3.185.

[16] Nicola Henry, Asher Flynn and Anastasia Powell (2018) 'Policing Image-Based Sexual Abuse: Stakeholder Perspectives' *Police Practice and Research*, 19(6), pp. 565-581, https://doi.org/10.1080/15614263.2018.1507892.

[17] Ajder, Henry, Giorgio Patrini, Francesco Cavalli and Laurence Cullen (September 2019). The State of Deepfakes: Landscape, Threats, and Impact, https://deeptracelabs.com/mapping-the-deepfake-landscape/.