UTS

CENTRE FOR MEDIA TRANSITION

Online Safety Amendment (Digital Duty of Care) Bill 2026

Submission from UTS Centre for Media Transition to the Senate Environment and Communications Legislation Committee

Date: 30 October 2026

About the Centre for Media Transition

The Centre (CMT) was established in 2017 as an applied research unit based at the University of Technology Sydney (UTS). It is an interdisciplinary initiative of the Faculty of Arts and Social Sciences and the Faculty of Law, sitting at the intersection of media, journalism, technology, ethics, regulation and business.

Working with industry, academia, government and others, the CMT aims to understand media transition and digital disruption, with a view to recommending legal reform and other measures that promote the public interest. In addition, the CMT aims to assist news media to adapt for a digital environment, including by identifying potentially sustainable business models, develop suitable ethical and regulatory frameworks for a fast-changing digital ecosystem, foster quality journalism, and develop a diverse media environment that embraces local/regional, international and transnational issues and debate.

This submission was prepared by:

- Dr Michael Davis, Research Fellow, UTS Centre for Media Transition

CONTACT

Centre for Media Transition
Faculty of Law, University of Technology Sydney
Building 2, Level 15
UTS City Campus, Broadway
PO Box 123, Broadway NSW 2007

cmt@uts.edu.au
+61 2 9514 9669

cmt.uts.edu.au

Introduction

Thank you for the opportunity to contribute to this inquiry into the Online Safety Amendment (Digital Duty of Care) Bill 2026 (the bill). The bill represents a positive step towards increasing accountability for digital platforms by imposing a duty on platforms to take action to prevent online harms. We support the introduction of a single, legally enforceable statutory duty of care that would place a general obligation on digital platforms to address the risks and harms arising from their systems and processes.

However, the bill has several potentially significant shortcomings that may reduce its likely effectiveness and result in negative impacts. These shortcomings lie particularly in the lack of clarity over whether the duty is aimed at platform systems or at protection from harmful content.

Imposing a duty aimed primarily at protection from harmful content, however narrowly defined, introduces a risk of prior restraint over freedom of expression and undue imposition on privacy by imposing a de facto obligation of general monitoring.

The duty should be directed explicitly at addressing the contribution of platform systems to online harms. If the duty is explicitly directed at platform systems, it can cover a broader range of online harms without raising the risk of general monitoring of user content and other undue impositions on privacy and freedom of expression. Risk assessment and risk management obligations should therefore form the heart of the duty, with these balanced by the obligation to consider impacts on rights including privacy and freedom of expression. Providing users with procedural rights to dispute content takedown and other platform actions would provide added protection.

We address these issues in more detail below.

Recommendations:

1. The duty should be explicitly directed at addressing the contribution of platform systems to online harms rather than general protection from harmful material and conduct.
2. Impacts on privacy, freedom of expression and other rights should be included in the consideration of appropriate and effective measures in cl 26A.
3. Complaints processes should explicitly include appeals processes for content moderation and other actions taken by platforms. The government should move forward with the establishment of an external body for escalated digital platform complaints.

The scope and target of the duty

1. Online harms result from a combination of user content and conduct on the one hand and platform design on the other. Digital platforms are vectors for harmful content and conduct, and the extent to which they are such vectors is in large part a function of platform design (such as content-recommender systems) and governance processes (such as content moderation, reporting systems and user terms of service).
2. There is increasing recognition that regulation of digital platforms for the purpose of reducing online harms should focus on platform systems rather than content. This focuses regulation on the aspects of online harms for which platforms bear clear responsibility. The report on the Statutory Review of the Online Safety Act (the Rickard review), observed that “online service providers exercise control over, and have knowledge of, their systems and processes, and must be held responsible for

the safety of their services".[1] A systems approach moves the regulatory focus from a reactive, ex-post scheme – largely enacted through the removal of harmful content – towards a preventive, ex-ante, approach. Aside from assigning responsibility appropriately, a further rationale for this approach is that it will better address the problems of scale that beset efforts to mitigate online harms.

3. A duty of care can be the foundation for a systems approach. However, as we noted in our submission to the Rickard review, it must be appropriately designed.[2] While there is an essential place for content takedown and moderation, particularly with respect to illegal and seriously harmful content, care must be taken that the duty does not require or create an incentive for platforms to engage in the generalised monitoring, proactive blocking or filtering of user content and activity. This is generally held to be an undue imposition on the rights to privacy and to freedom of expression and shifts too much of the responsibility for online harms away from user conduct and onto platforms. Comparable international platform regulation includes safeguards against such impositions, for example the prohibition on general monitoring in the EU Digital Services Act (DSA) and protections for freedom of expression in the UK Online Safety Act (UK OSA).

4. The explanatory memorandum (EM) suggests that the intention of the bill is to focus on platform systems and processes. On the face of the bill, however, this intention is less clear. The duty is not explicitly focused on the contribution of platform systems but instead on the responsibility to ensure a safe online environment, which is in turn defined in terms of protection from specific classes of harmful material and conduct. Taken on its own, the duty could incentivise platforms to proactively block or filter content, and to block or filter more than is required by the bill in order to assure compliance. Further, the bill does not rule out general monitoring and there are no explicit protections for privacy and freedom of expression. Hence, this incentive is not counterbalanced by protections for user rights.

5. By tying the duty to protection from specific types of harmful content and behaviour, there is also a danger that the duty will fail to adequately address the contribution of platform systems and processes to online harms. Providers could reasonably argue that the most reasonable, appropriate or effective means of ensuring a safe online environment is by removing harmful content. This would effectively undermine a fundamental purpose of introducing the duty.

6. The explanatory memorandum (EM) observes that the intent of the bill is for providers to implement and maintain effective systems and processes to mitigate the risk of harm. While "individual pieces of harmful content found on an online service will not necessarily constitute a breach of the Duty … the presence of harmful content may form part of the evidence that there is a systemic issue constituting a breach" (p.67). However, the duty as worded in cl 26 is "to ensure a safe online environment", where this is defined as protection from harmful material and conduct. Hence on the face of the bill, the presence of harmful material and conduct is not merely part of the evidence of a systemic breach, but the very focus of the duty.

7. The EM notes in this context that establishing a test of what measures are appropriate is important in setting the focus on platform systems. We agree; however, this test is also defined in terms of whether the measures would achieve a

[1] D Rickard, *Report of the Statutory Review of the Online Safety Act 2021*, October 2024, p. 50, <https://www.aph.gov.au/Parliamentary_Business/Tabled_Documents/9184>.
[2] M Davis et al., *Statutory Review of the Online Safety Act 2021, Submission from UTS Centre for Media Transition to the Department of Infrastructure, Transport, Regional Development, Communications and the Arts* (Submission 134), 28 June 2024, p. 9.

safe online environment, which, as we have noted, is defined in terms of protection from specific classes of harmful material and conduct.

8. Similarly, the Carnegie Trust has argued, with respect to the UK OSA, that "the fact that the threshold for action is defined by specified types of content may make it difficult to assess the contribution of the systems design on content creation and dissemination."[3]

9. As noted in the EM, the list of seriously harmful content and conduct in cl 25D is intended to be exhaustive and closely aligned with what would be considered illegal or class 1A content under the online content scheme. But, as observed by Woods and Perrin in their work recommending a digital duty of care in the UK, "the criminal law does not constitute a complete list of harms against which we would expect a service provider to take action."[4] The danger of focusing on a narrow and exhaustive set of harms is that the incentive for platforms to improve their general systems is greatly reduced.

10. Clause 27 requires platforms to undertake risk assessments "for the purposes of complying with their digital duty of care". Again, given the definition of the duty, this leaves risk assessments tied to the narrow and exhaustive lists of serious harms for adults and harms to children.

11. Directing the duty explicitly at improving platform systems would also allow the duty and due diligence obligations to encompass a broader range of online harms than is currently contemplated by reducing incentives for platform overreach or prior restraint.

12. The consideration of appropriate and effective measures in cl 26A requires the proportionality of those measures to be taken into account. While assessing the impact on human rights necessarily requires an assessment of proportionality, the converse is not necessarily true. To ensure that protection from harm is balanced with respect for rights including privacy and freedom of expression, consideration of these and other rights should be included in the consideration of whether provider actions and measures are appropriate and effective.

Recommendation 1: The duty should be explicitly directed at addressing the contribution of platform systems to online harms rather than general protection from harmful material and conduct.

Recommendation 2: Impacts on privacy, freedom of expression and other rights should be included in the consideration of appropriate and effective measures in cl 26A.

Complaint and dispute mechanisms

13. The actions taken by online service providers impact the rights and interests of their users. An important element in promoting the accountability of providers is ensuring that users' rights and interests are protected through the provision of robust complaints mechanisms. Given that the digital duty is designed to create a "stronger and proactive regulatory framework", it is likely, as we noted, to incentivise over-broad content takedown and moderation. Hence, it is critical that users have access to dispute mechanisms that mitigate this problem by, for

[3] Carnegie UK Trust, *The Draft Online Safety Bill: Carnegie UK Trust Initial Analysis* (2021), para. 27, https://carnegieuk.org/blog/the-draft-online-safety-bill-carnegie-uk-trust-initial-analysis/, accessed 27 Oct. 2026.

[4] Lorna Woods and William Perrin, *Online Harm Reduction – a Statutory Duty of Care and Regulator* (Carnegie UK Trust, 2019), p. 40, https://doi.org/10.2139/ssrn.4003986.

example, reinstating content that has been taken down even though it does not violate the provider's terms of service or constitute a harm under cl 25.

14. Currently, however, the bill requires only that platforms have complaint and dispute processes. While the Commissioner may determine further requirements for those processes, we note that the objects of the Act at clause 3(c) do not refer to procedural rights but only to "providing complaints schemes or mechanisms to quickly investigate, remove and remediate illegal or seriously harmful material."
15. In our submission to the Rickard review, we also argued that an external dispute resolution mechanism should be established to promote provider accountability. This is essential for ensuring that users have recourse where they are not satisfied with the outcome of an internal complaint mechanism. We also recommended that it take the form of an ombuds scheme. We cover the rationale and potential models for an external complaints scheme in our publication, *Digital Platform Complaint Handling - Options for an External Dispute Resolution Scheme*.[5]

Recommendation 3: Complaints processes should explicitly include appeals processes for content moderation and other actions taken by platforms. The government should move forward with the establishment of an external body, such as an ombuds scheme, for escalated digital platform complaints.

[5] Holly Raiche, Derek Wilding, Karen Lee & Anita Stuhmcke, *Digital Platform Complaint Handling: Options for an External Dispute Resolution Scheme* (UTS Centre for Media Transition, 2022). DOI:10.26195/7m5k-wp38