

IPA RESEARCH NOTE

Five reasons the Digital Duty of Care will enable government censorship of the internet

SEPTEMBER 2026

Dr Andrew Bushnell

Director, Law and Freedom Program

Key findings

The federal government's proposed Digital Duty of Care is conceptually flawed and will necessarily impinge on Australians' rights to free speech, association, and political participation.

IPA analysis of the available government materials outlining the proposal identifies five dangers inherent in the Digital Duty of Care:

- The meaning of “serious harms” is vague, subjective, and will expand over time.
- The duty to eliminate intangible harms will allow the government to smuggle its abandoned misinformation laws in through the backdoor.
- The wide and uncertain scope of the duty will incentivise over-compliance by regulated companies.
- Compliance with the Digital Duty of Care will require increased surveillance of internet users.
- The duty will be enforced by the unaccountable Office of the eSafety Commissioner.

Background

The Digital Duty of Care is intended to impose an “overarching” requirement on all online service providers to “provide a safe online environment for all Australians” by taking “reasonable steps to maintain effective systems and processes” that “manage risks of serious harm associated with the use of their service[s]”. Online service providers would include not just social media platforms, but other services such as messaging applications, and the duty would be enforced by the Office of the eSafety Commissioner, which presently administers the *Online Safety Act 2021* and associated regulations.¹

The federal government committed to the Digital Duty of Care in November 2024. This followed the final report of a statutory review of the *Online Safety Act 2021*, which, along with recommending a “singular and overarching duty of care”, made 67 recommendations, including that the Commonwealth strengthen the eSafety regime through new powers, stronger penalties, and changes to its governance structure.² The Digital Duty of Care was among the review's 41 recommendations the government gave full or partial support to, along with a further 23 recommendations noted for further consideration.³

While the federal government is yet to release a draft bill for the Digital Duty of Care, the review, government response, and later issues paper provide sufficient preliminary detail of what the duty is and what it may entail in practice – and together make plain the risk the Digital Duty of Care carries for freedoms online.

1. The meaning of “serious harm” is vague and will expand over time

The Digital Duty of Care poses an inherent risk of legal, non-harmful speech and information being censored, simply because the duty covers a non-specific and changeable set of harms.

The concept is premised on a misleading analogy between internet regulation and regulation pertaining to physical safety, like workplace health and safety law or trade practices law.⁴ This approach misses what is distinctive, and most important, about regulating the internet: the object of the regulation is not a physical environment, but a place shaped by, and used for, the thought, expression, and interactions of human minds.

To the extent that the internet causes physical harms, it is always at a distance: someone must be influenced by what is said online and then choose to act upon it in real life. Otherwise, the harms of the internet are chiefly psychological and mental. This means that compliance with the duty is always fraught with contested political and moral questions to do with the kinds of ideas and beliefs that are healthy for humans – subjects that make up much of political discourse in a free society.

Because the duty seeks to control this wide-ranging and contested space, the harms it covers are necessarily vague and changeable.

The Digital Duty of Care anticipates a wide range of non-specific harms

The Digital Duty of Care aims to make online service providers responsible for protecting users against “reasonably foreseeable serious harm”.⁵ Accordingly, the federal government is considering a wide range of non-specific harms that would give the Minister and eSafety considerable discretion in both defining harm and identifying new harms to regulate.

For example, in a speech in November 2024, then–Communications Minister Michelle Rowland, adopting language from the statutory review, stated that the Digital Duty of Care legislation will include “enduring categories of harm [such as:] harms to young people; harms to mental wellbeing; the instruction and promotion of harmful practices; and other illegal content, conduct or activity”.⁶ The statutory review itself recommends that the legislative harms ought to include, “at a minimum”, a wide range including both serious (and already illegal) behaviours as well as less concrete notions like “problematic internet use” and “threats to... social cohesion”.⁷

These are vague and imprecise terms. Such ambiguous category descriptions could be abused, for example, to censor a social media account critical of the federal government’s migration policy under the auspices that it poses a threat to “social cohesion”. Nor is there any social consensus around what constitutes “harms to young people”, “harms to mental wellbeing”, and “harmful practices”.

An early indication of how the duty will likely be interpreted can be seen in the federal government’s response to the report of its Special Envoy to Combat Islamophobia, which points to its commitment to the Digital Duty of Care as part of its response to “combatting online hate”.⁸

The Minister and the eSafety Commissioner will have the power to expand the categories of harms

The government’s issues paper specifies that the Minister for Communications and, at the Minister’s direction, the eSafety Commissioner will have the power to make “binding rules” for “designating additional categories and types of harm” and “specifying compliance requirements for service providers”.

Regarding protections, the issues paper notes (as did the earlier government response to the review) that these rules will be subject to parliamentary oversight, and the federal government’s response notes that the legislation will “focus on protecting vulnerable Australians including children, combating criminal activity and promoting safety, while respecting Australians’ freedom of expression”.⁹

But broad categories of “enduring harms” in the legislation will give the Minister considerable flexibility – this would be needed to give full effect to the intention of the scheme to cover all foreseeable harms across the entire internet, and the expansiveness of the duty means that balancing the demands of “safety” with Australians’ rights and freedom will inevitably favour the former at the expense of the latter.

“Serious” does not adequately limit the scope of the harms covered by the duty

The modifier *serious* does not provide meaningful protection for Australians nor limit the scope of the duty, as this too will be a matter of interpretation for the regulator.¹⁰

The *Online Safety Act 2021* defines serious harm as “serious physical harm or serious harm to a person’s mental health”, with the latter further defined as serious psychological harm or serious distress excluding “mere ordinary emotional reactions”.¹¹ This definition is used in the regulation of cyber abuse material targeted at adults but nowhere else in the Act. In 2024-25, eSafety reported that approximately 10 per cent of the 3,600 complaints of adult cyber abuse it received met the serious harm threshold.¹²

However, it has also been revealed that eSafety issues informal takedown notices even for material that does not meet this threshold, as seen in *eSafety Commissioner v Baumgarten* [2026], highlighting that even where there is a supposedly high statutory threshold, the effect of legislating against harm has a wider regulatory effect.

2. The duty to manage risks of harm will smuggle in a more expansive version of the abandoned misinformation/disinformation laws

The proposed Digital Duty of Care appears to be a stealthy and steroidal version of the federal government’s failed 2024 bill to regulate so-called misinformation and disinformation online.

The vague concept of “harm” provides a direct connection between the two proposed attempts to police speech on the internet. The misinformation/disinformation draft bill purported to restrict the dissemination of information that was “false, misleading, and deceptive” and “likely to cause or contribute to serious harm” (whether done with or without intent to deceive). In that bill, “harm” was defined broadly to include “hatred against protected groups; disruption of public order or society; harm to Australian democratic processes or government institutions; harm to Australians’ health; environmental harm; and economic or financial harm”.¹³

As noted, the Digital Duty of Care uses the same regulatory architecture, built around the concept of harm. The issues paper makes clear that the duty is not intended to be limited to illegal content but intended to cover harmful content and design features of online services.¹⁴ The statutory review argues for “a comprehensive approach to online harms”. It cites the regulatory frameworks of both the European Union and United Kingdom as examples of the duty of care systems, noting that those jurisdictions target misinformation and disinformation as “societal harms”, in “diluting the availability of public interest journalism”, and “national security harms” such as misinformation and disinformation during an election. Therefore, there is nothing inherent in the Digital Duty of Care that would exclude the earlier categories of misinformation/disinformation harms from its scope.

Moreover, it is likely that the Albanese government intends for those kinds of harm to be covered by the duty. The Albanese government announced its commitment to introducing a Digital Duty of Care in November 2024.¹⁵ This was the same month that it was forced to withdraw the earlier bill that would have given the Australian Communication and Media Authority power to regulate misinformation and disinformation online. At the time, then-Communications Minister Michelle Rowland presented the Digital Duty of Care as complementary to the earlier bill, as two parts of a broader strategy.¹⁶

At the time of the earlier bill, IPA research found that the concept of harm was subjective and vague and would “turn big tech companies into the censorship enforcement arms of the state”.¹⁷ IPA research also found that Australians reject these attempts to censor the internet.¹⁸

These concerns are even weightier in respect to the Digital Duty of Care. The Digital Duty of Care will not be limited to information that is false, misleading or deceptive, but will cover all information on the internet. The duty is also to be implemented through opaque government influence on the design of online services. These so-called “safety by design” principles are discussed below.

The harms that were intended to be captured by the earlier bill are easily recast as foreseeable harms within the scope of the duty. For example, during the Covid-19 vaccine rollout, sitting MP Craig Kelly’s social media accounts were taken down at the request of the federal government on the basis that they were spreading misinformation. Under the proposed Digital Duty of Care, this might be recast as serious harm (to users who might refuse the vaccine, to public health, or to public institutions), and similar accounts might be taken down or, indeed, topics of conversation might be censored wholesale, with prohibitions built into platforms and their algorithms.

Just as the *Online Safety Act 2021* uses the term “cyber abuse” to describe harsh words and online bullying in a way that invites confusion with sexual abuse, the Digital Duty of Care obscures its aim of censorship with the language of harm.

3. The wide scope of the duty will incentivise over-compliance and risk aversion among regulated companies

Another way in which the Digital Duty of Care is a sprawling concept is in the breadth of its application across the internet. The duty is intended to cover platforms and services at multiple levels of the “tech stack” (as the issues paper describes it).

The duty treats the internet as a set of interconnected and potentially harmful systems. In the words of the issues paper, the duty will apply to “all online service providers currently defined as within the scope of the Act [their emphasis]”. To take just some of the issues paper’s examples, these services will include: social media services; messaging apps; interactive online games; and online dating services; many generative AI services (including chatbots and image generators); search engine services; app stores; and operating system services [for] phones, tablets, laptops, wearable devices.¹⁹

Although the government claims the duty will apply to different companies differently depending on their risk profiles, it is ultimately a simplistic, one-size-fits-all approach that will incentivise online service providers to implement a restrictive interpretation of the duty and, in turn, limit Australians’ ability to use the internet to access information and to communicate with one another.

Regulated companies will converge on a risk-averse approach

To satisfy the regulator, regulated companies will likely converge on a common risk assessment based on the highest-risk company. Because these regulated companies’ businesses interact – for example, app stores host social media apps, games, and AI tools, and all are enabled by devices’ various operating systems – the easiest way to acquit their various duties is to establish common standards. Operating systems may simply not allow applications that are deemed high-risk to be installed, social media sites may not allow AI-enabled posts, search engines may not index sites, and app stores may not host games, and so on.

Additionally, compliance with this standard will not be equally achievable for all online service providers; the costs will likely lead to less diversity in services across the internet.

In recent consultations, stakeholders have already flagged the need to “avoid cumulative as well as duplicative burden” and a preference for the United Kingdom’s model of legislating multiple specific duties rather than the European Union’s model, which is closer to that proposed by the government.²⁰

The duty promotes over-compliance among regulated companies

With new powers, strong penalties, and a wide engagement remit, in practice, the Digital Duty of Care will lead to most online service providers over-complying with the law. Just as they are likely to converge on risk averse common standards, they are also likely to create systems and processes that capture more content than is legally required.

The federal government has supported strengthening the enforcement powers of eSafety. As noted, its recent bill on the under-16 social media ban would give new investigatory powers to the regulator. For the Digital Duty of Care, it also supports giving the regulator new powers to “use enforceable undertakings or issue remedial directions to services in relation to all relevant penalty provisions” – meaning that in addition to civil penalties, eSafety would be able to order companies to take actions to remove content or remediate systems before escalating to penalties for non-compliance.

Regarding penalties, along with its recent doubling of fines for social media companies found not to be complying with the under-16 social media ban, the Albanese government has also flagged stronger civil penalties for companies failing to implement the Digital Duty of Care. The issues paper suggests that the maximum civil penalty for non-compliance should be \$100 million.²¹ But the government has supported a larger set of penalty recommendations coming from the statutory review, including \$10 million penalties for non-compliance with content removal notices, and a penalty of up to 5 per cent of companies’ global annual turnover.²²

In addition, as noted in previous IPA research on the defeated misinformation/disinformation bill, because government will have outsourced censorship to private entities, Australians will have little to no legal recourse when lawful content is swept up in the state-mandated dragnet.²³

“Safety by design” will give eSafety extraordinary reach inside companies

Online service providers will be required to not only police certain kinds of harmful content, but also to adopt principles of “safety by design”. This idea follows from the concept of an overarching duty. The statutory review argues that a single duty is preferable to multiple duties aimed at specific kinds of undesirable content, in part, because to comply, online service providers will both have to restrict harmful content and design systems that avoid causing harms through their operations (by creating “dark patterns” of behaviour like excessive use or rabbit holes).²⁴

In practice, safety by design means “assessing and addressing online harms in the design and provision of... services,” along with requirements for considering the interests and rights of users and accountability for “innovation and sharing of good practice in online safety”. This may include building in measures that prevent or remove seriously harmful content, preventing AI tools from generating such content, and designing algorithms to suppress the spread of such content. But beyond content, the idea is also to limit features that enable harmful user behaviours, like extended use or permitting anonymous/pseudonymous accounts.

By bringing safety by design principles within eSafety’s remit, the Digital Duty of Care gives the regulator unprecedented influence over online services, from the inception and development to the evaluation of services. The issues paper even notes that “eSafety has proven expertise and experience in providing research and guidance on how industry can embed safety by design principles”.²⁵

The implication is that eSafety’s remit will include advising online service providers on how to build their applications. Compliance and reporting may not be enough; full cooperation with eSafety’s directions will be expected. This is already the norm in the European Union, which has recently moved towards ordering Meta to disable “infinite scroll” and other features on its platforms.²⁶

Companies will have internal compliance teams to align with government edicts and preferences

Not only will eSafety have a say in how services are designed, but it will also, in effect, monitor companies from the inside. One of the recommendations of the statutory review, which was accepted in principle by the Government, was that services with wide reach or high risk should have “well-resourced” compliance teams that report directly to senior management and the board, and that the head of compliance should serve at the board’s pleasure.²⁷

The function of mandating this kind of internal oversight is to ensure that companies are aligned with the expectations of government. The jobs of the compliance officers will be to continually engage with Government to understand compliance requirements and, internally, to oversee content moderation systems and content reporting. If adopted, this policy would be akin to Government installing cells of political apparatchiks within companies – a dynamic familiar in authoritarian states but (relatively) novel in Australia.

4. Compliance with the Digital Duty of Care will require increased surveillance of internet users

Just as online service providers will be under government surveillance, to meet their obligations for managing content and reporting compliance, they will need to increase their surveillance of users.

Under the duty, internet companies will be obliged to conduct regular risk assessments and evaluations of their risk management systems. This will be backed by reporting obligations and a power for eSafety to order companies to undertake independent audits, at their own expense, of their compliance systems and processes.²⁸

To properly assess these risks and report on their incidences, online service providers will need to know more about users – their demographics, profiles, preferences, and habits. Tracking this is already a part of many companies’ business models but is driven by the incentive to make applications more attractive to users. Overlaid on this now will be an imperative to make sure that users conform to government expectations in their overall patterns of behaviour. In addition, this user information would have to be shared with eSafety (and, likely, with “authorised independent researchers” – an opportunity for doxxing accounts disfavoured by government).²⁹

This will have both a direct and an indirect effect on users’ freedoms of speech and association. The direct effect will come from platforms’ policing (likely also to be highly automated) of users’ activities and patterns. The duty will also diminish the possibility that services and their user communities might develop their own behaviours and expectations – that, say, cater to niche preferences – with a consequent loss of internet diversity. Indirectly, there is likely to be a chilling effect on speech as users reduce their use of a range of online services, like social media platforms, as they reassess the privacy trade-offs involved in their use.

5. The Digital Duty of Care will be enforced by an unaccountable Office of the eSafety Commissioner

Lastly, all this power will be administered by eSafety, a regulator that has a questionable record of accountability and politicisation.

The formal compliance requirements of the duty are onerous – and they are also unlikely to be the full extent of the regulatory effect on online service providers. IPA research has detailed the informal practices of eSafety in relation to adult cyber abuse: the regulator routinely contacts platforms about taking down content, even when the content does not meet the legal threshold. This practice was exposed in the *Baumgarten* case, in which the Federal Court found that eSafety had been exceeding its powers by presenting informal notices as though they were backed by legal power.³⁰ By broadening and strengthening eSafety’s enforcement powers,

the Digital Duty of Care will also increase its rate of engagement with the domain, leading to more informal interventions by the regulator and likely more unlawful actions, all of which will be obscure to the public.

The politically fraught nature of the eSafety office can be seen in the decision to exclude the social media platform BlueSky from the under-16s social media ban, while including similar platforms like X (formerly Twitter). The regulator claimed this was because the assessed risk of BlueSky was lower because of the small base of Australian users on the platform. But this only raises questions about how eSafety weighs factors like reach and content. It is difficult to avoid the conclusion that the content on BlueSky was considered less weighty because the political tendencies of that site are considered less “harmful” than similarly political content on X.

Beyond this, the eSafety Commissioner has also taken a broad view of the office’s public advocacy role, becoming a prominent commentator on issues outside the office’s remit, including the proposed misinformation/disinformation laws. The eSafety website also provides information that goes well beyond protecting people from unlawful behaviour, like, for example, propaganda aimed at “young men and boys” about “toxic” masculinity.³¹

Yet despite all this, the Albanese government intends, by implementing the Digital Duty of Care, to give eSafety both greater reach and greater resources. The statutory review recommended replacing the statutory office of the Commissioner with a multi-member commission, eventually establishing it as a standalone agency with its own legal team and corporate resources, potentially funded by industry.³² In its response, the government supported in principle the commission model and supported additional resourcing, but has so far only noted the proposals for a standalone agency with dedicated funding streams.³³ But even without those further changes, the proposed duty would substantially enlarge eSafety’s operational reach through rule-making at the Minister’s direction, expanded information-gathering and investigatory powers, compulsory audits, and remedial directions.³⁴

As a regulator, eSafety already frequently abuses its powers. It cannot be trusted with the kinds of powers that an overarching duty of care necessarily entails.

For more information contact Morgan Begg, Director of Research at mbegg@ipa.org.au

End Notes

1. Department of Infrastructure, Transport, Regional Development, Communications, Sports and the Arts, *A Digital Duty of Care for Australia – Developing a duty of care framework for online services used by Australians* (Issues Paper, May 2026) (“Issues Paper”): <https://www.infrastructure.gov.au/department/media/publications/digital-duty-care-australia-developing-duty-care-framework-online-services-used-australians>.
2. Delia Rickard, *Report of the Statutory Review of the Online Safety Act 2021* (Commonwealth Government, October 2024): <https://www.infrastructure.gov.au/have-your-say/statutory-review-online-safety-act-2021>.
3. Commonwealth Government, *Australian Government Response to the Independent Review of the Online Safety Act 2021* (April 2026) (Government Response”): <https://www.infrastructure.gov.au/department/media/publications/australian-government-response-independent-review-online-safety-act-2021>.
4. Statutory Review, 7; Michelle Rowland, “Speech – The Sydney Institute – The Governance of Digital Platforms” (13 November 2024) (“Rowland 2024a”): <https://minister.infrastructure.gov.au/rowland/speech/speech-sydney-institute-governance-digital-platforms>.
5. Issues Paper, 2.
6. Rowland (2024a); The language of “enduring harms” comes from the Statutory Review, 54.
7. Statutory Review, Recommendation 5.
8. Commonwealth Government, *Combating Islamophobia: Australian Government Response to the Special Envoy’s National Response to Islamophobia* (July 2026): <https://www.homeaffairs.gov.au/reports-and-pubs/PDFs/ag-response-special-envoy-combat-islamophobia.pdf>.
9. Government Response, 3.
10. This point was noted by stakeholders consulted after the release of the Issues Paper. Commonwealth Government. *Digital Duty of Care Stakeholder Consultation May-June 2026: Key Outcomes* (July 2024): <https://www.infrastructure.gov.au/sites/default/files/documents/digital-duty-of-care-summary-of-consultation-outcomes-2026.pdf>
11. *Online Safety Act 2021* (Cth) s 5.
12. Australian Communications and Media Authority, *Annual Report 2024-25* (2025) 210: <https://www.esafety.gov.au/sites/default/files/2025-10/ACMA-eSafety-annual-report-2024-25.pdf?v=1784074162859>.
13. *Exposure Draft: Communications Legislation Amendment (Combating Misinformation and Disinformation) Bill 2023*. <https://www.infrastructure.gov.au/sites/default/files/documents/communications-legislation-amendment-combating-misinformation-and-disinformation-bill2023-june2023.pdf>.
14. Issues Paper, 5,8
15. Michelle Rowland, “New Duty of Care Obligations on Platforms will Keep Australians Safer Online,” (Media Release, 14 November 2024) (“Rowland 2024b”): <https://minister.infrastructure.gov.au/rowland/media-release/new-duty-care-obligations-platforms-will-keep-australians-safer-online>.
16. Rowland (2024a).
17. John Storey and Morgan Begg, *Submission to the Inquiry into the Communications Legislation Amendment (Combating Misinformation and Disinformation) Bill 2024* (Institute of Public Affairs, 2024): <https://ipa.org.au/read/institute-of-public-affairs-submission-to-the-inquiry-into-the-misinformation-and-disinformation-bill>.
18. Daniel Wild, *Poll: Australians Reject the Federal Government’s Online Censorship Bill* (Media Release, Institute of Public Affairs, 23 October 2024): <https://ipa.org.au/read/poll-australians-reject-the-federal-governments-online-censorship-bill>.
19. Issues Paper, 2-3.
20. Stakeholder Consultation, 2.
21. Issues Paper, 7.
22. Statutory Review, 137-8; Government Response, 11.
23. Morgan Begg and John Storey. *Canberra’s Digital Ministry of Truth* (Institute of Public Affairs, August 2023) 13: <https://ipa.org.au/read/canberras-digital-ministry-of-truth>
24. Statutory Review, 53.
25. Issues Paper, 4.
26. Ana-Maria Stanciuc, “EU tells Meta to dismantle Facebook and Instagram’s ‘addictive’ design or face a fine” (*The Next Web*, 10 July 2026): <https://thenextweb.com/news/eu-orders-meta-addictive-design-changes>.
27. Statutory Review, 67.
28. Issues Paper, 6-7.
29. Issues Paper, 7. Statutory Review, 69.
30. Margaret Chambers, “A power unto herself – Part one” (*Looking Forward*. 1 June 2026): <https://theipa.substack.com/p/a-power-unto-herself-part-one>.
31. eSafety. “Young men and boys”. <https://www.esafety.gov.au/key-topics/young-men-and-boys>.
32. Statutory Review, 24-25, 169-181.
33. Government Response, 14-15.
34. Issues Paper, 3,6,7.