Policy Submission

Submission to the Review of Model Defamation Provisions - Stage 2

To: *NSW Attorney-General*
From: *Reset Australia*

Reset Australia is an independent, non-partisan organisation committed to driving public policy advocacy, research, and civic engagement agendas to strengthen our democracy within the context of technology. We are the Australian affiliate of Reset, the global initiative working to counter digital threats to democracy. As the Australian partner in Reset's international network, we bring a diversity of new ideas home and provide Australian thought-leaders access to a global stage.

We look forward to working through this consultation and beyond, as we push this conversation forward to ensure appropriate and considered legislation that protects Australian institutions, citizens and democracy.

Matt Nguyen

Contact: hello@au.reset.tech

Reset.
AUSTRALIA

CONTEXT

Internet companies, and in particular digital platforms and search engines, have revolutionised our collective relationship with information. At no other point in history have people been able to create and disseminate content to as large a potential audience without the need to engage with traditional gatekeepers such as publishers or news organisations. Whilst this democratisation of access should be celebrated, it has also added complexity and nuance to our existing regulatory infrastructure.

The business models of the digital platforms have a single objective - to capture and maintain user attention in order to maximise advertisements served and profits generated. As such, the algorithms which dictate the content and information we consume are optimised to fulfil this objective, resulting in an attention economy. To feed this machine, the platforms have built a sophisticated system of unfettered personal data collection, building comprehensive profiles of their users that encapsulate their interests, vices, political leanings, triggers and vulnerabilities. This data is then used to predict our engagement behaviour, constantly calculating what content has the greatest potential for keeping us engaged. This content has been shown to lean towards the extreme and sensational, as it is more likely to earn higher engagement[1],[2].

This has resulted in the explosion of a data economy that has been facilitated through the commoditisation of personal information. This model, termed 'surveillance capitalism' by Shoshanna Zuboff,[3] is predicated on the extraction and exploitation of personal data for the primary purpose of predicting and changing individual behaviour. This emerging model (spearheaded by Google and later Facebook) sets a dangerous precedent for adoption by other industries, and flies against Australian ideals of autonomy, public safety and privacy.

The determination of online intermediary liability is one of the most contested topics within the broader conversation of digital platform regulation and extends beyond defamation, however it is within this context of the attention economy monopoly that the major digital platforms have created that any policy development must occur. From online harm to copyright, there is an ongoing and fragmented legal debate on why, how and when we might determine the degree of liability these companies should hold - and whilst defamation is the focus of this review, a harmonised approach across State and Federal) must be pursued to ensure an appropriate new regulatory scheme can be developed to address these issues.

---

[1] Vosoughi et al. (2018), 'The spread of true and false news online', *Science* found at https://science.sciencemag.org/content/359/6380/1146

[2] Nicas (2 Feb 2018), 'How YouTube Drives People to the Internet's Darkest Corners', *Wall Street Journal* found at https://www.wsj.com/articles/how-youtube-drives-viewers-to-the-internets-darkest-corners-1518020478

[3] Zuboff S (2019), 'The Age of Surveillance Capitalism,' Profile Books, London

POLICY APPROACH

The focus of this submission will be on the role of digital platforms, in particular content aggregation services, social media and instant messaging services.

Striking a balanced and nuanced approach to intermediary liability, particularly within the context of defamation is difficult. Many critics are quick to point out the tensions and trade offs that a heavy-handed approach might cause ranging from undue compliance burdens to placing undue limits on freedom of expression.

In a submission to this review by Digi (Australia's Big Tech industry lobby group), they state:

> *Internet intermediaries, including social media websites and search engines, are not the creators of defamatory content, and do not have the ability to determine whether any allegedly defamatory content is true or would otherwise be defensible.*

Whilst we agree that these intermediaries do not create content, their impact in its dissemination is arguably greater than any individual user. If defamation law is to provide redress for reputational damage, it's in this amplification (that is so fundamental to these platforms' business models) that much of the harm occurs. Whilst we sympathise with the need to afford certain protections for intermediaries, too often the 'all or nothing' approach illustrates that these companies are all too ready to shirk responsibility and not be held accountable.

This balance between outright liability - which we agree could pose serious threats to freedom of expression - and the clear need for more accountability on the part of the platforms. Pappalardo and Suzor express this as the tension between the principle that there is no right without remedy and the principle that there is no liability without fault[4]. They go on to suggest a refocus to prioritising causal responsibility in the evaluation of online intermediary liability as a potential pathway for greater clarity.

Our recommendations build on this analysis and provide a pathway to ensure that:

1) there is a pathway for remedy and,
2) there are more transparent mechanisms to determine fault

This can only be achieved through legislative measures that compel the digital platforms to be more open with their data and processes, and accountable to the public interest.

---

[4] Pappalardo, Kylie & Suzor, Nicolas (2018) The liability of Australian online intermediaries. *The Sydney Law Review,* 40(4), pp. 469-498.

RECOMMENDATIONS

1) A centralised and open complaints process which prioritises private mediation

Ensuring that there is an open and accountable process in which defamation complaints can be raised and dealt with transparently. This must be managed and held by a publicly accountable body (similar to an ombudsman or a complaints facility within an independent regulator) so that individuals have a dedicated facility for recourse. Whilst failure to deal with complaints from the intermediary side opens them up to liability, the current process puts an undue burden on individuals.

Additionally, digital platforms should be compelled to openly publish and publicise their complaints processes and de-identified results.

> **Facilitated Platform Mediation**
> A possible mechanism to promote the private mediation of defamation whilst relieving the burden on platform companies is an automated mediation process, wherein alleged posters of defamatory material are notified (and given options to take down or seek in-person mediation) when a complaint has been made against them.

2) Transparency and investigative powers

A significant barrier to assessing liability is the inability to effectively demonstrate causal responsibility in the evaluation of intermediary actions. As such, an independent body must be given mandatory investigative powers via algorithmic audits.

The systematic impacts of algorithmic amplification - that is the promotion/demotion of content that is currently dictated by the digital platform's internal algorithmic processes - is an issue that goes far beyond traffic and advertising revenue, and requires an expansive remit to address. Unilateral algorithmic curation and amplification has an outsized harmful impact on the impact of defamatory material.

This information is held solely by the digital platforms, who do not make it available for transparent independent review under any circumstances. The digital platform companies have all the data and tools needed to understand their role in disseminating defamatory material. Without mandated access, we are forced to rely on the companies to police themselves through their own internal policies.

Algorithmic Audits
An algorithmic audit is a review process by which the outputs of algorithmic systems (in this case the curation systems of the digital platforms which display content) can be assessed for unfavourable, unwanted and/or harmful results. In addition to assessing if design decisions within the digital platform algorithms were actively made that contribute to the dissemination of defamatory material, this process can also be expanded to examine organisation's internal processes, review points and decisions which may lead to any assertions of culpability.

How would an audit authority work?
The authority must have the ability to carry out an algorithm inspection with the consent of the digital platform company; or if the company does not provide consent, and there are reasonable grounds to suspect they are failing to comply with requirements, to use compulsory audit powers. It must be resourced (financially and technically) to carry out these actions, but it should also have the power to instruct independent third-party experts to undertake an audit on their behalf.

3) Develop a doctrine of accountability to complement proportional liability

One of the core questions of this review is how might you compel digital platforms to be more 'responsible' without relying on strict liability. Incentivising measures such as publishing public notices or issuing warnings that rely on diminishing organisational reputation prove wholly ineffective due to the market dominance of these platforms.

Ultimately, responsibility must be compelled through the expectation of users that they deserve certain services and protections - especially when they feel that they have been wronged. Short of running consumer education campaigns to illustrate how users might pressure these organisations to take on greater responsibility, we propose some mechanisms which might create an enabling environment for more desired outcomes.

a) Proportionate designation of digital platforms

Have a separate classification (based on % active users within Australia) for digital platforms that have an outsized potential to facilitate defamatory harm. Having a separate class of remedies (much larger fines based on % turnover as an example) for when these 'outsized impact' digital platforms are found liable might serve as both greater deterrent and a signal to smaller platforms to ascribe to greater responsibility.

b) Develop an industry standard for internet intermediary companies to share processes and best practices around defamation

Ensuring industry accountability, through co-created standards could be another way to spur responsibility. This was the starting point for some other issues digital platforms face (such as content moderation and misinformation) and can provide a useful starting point to drive solutions and commitments to this complex policy area.

A potential model for replication could be the eSafety Comissioner's 'Safety by Design' guidelines. There is a grey space in which these platforms operate, where certain actions can make them more or less 'liable' - as such these companies should work together to develop best practices.

A 'Responsibility by Design' Code might see digital platforms committing to;

- Proactive outreach to potential users who have been subject to alleged defamation
- Transparent automated processes to pick up potentially defamatory material
- Dedicated internal capacity to deal with complaints, including appropriate collaboration with government and legal resources
- Automated shadow-ban / de-amplification tools that automatically 'hide' alleged defamatory material until proper assessment can be made