Policy, Reform & Legislation
Department of Communities and Justice
defamationreview@justice.nsw.gov.au

31 May 2021

QUT Digital Media Research Centre submission in response to the Stage 2 Review of the Model Defamation Provisions

Prepared by Lucinda Nelson, Prof Nicolas Suzor and Dr Kylie Pappalardo

We are researchers in QUT's Digital Media Research Centre and the ARC Centre of Excellence for Automated Decision-Making and Society. The DMRC is a global leader in digital humanities and social science research with a focus on communication, media, and the law. The ADM+S is a cross-disciplinary national research centre that supports the development of responsible, ethical and inclusive automated decision-making systems.

For more information about this submission, contact Prof Nicolas Suzor:

Executive summary

We welcome the opportunity to contribute to this important review. In general terms, we are concerned with the operation of Australian defamation law in practice. While relatively few cases make it to the courts, **current law creates dangerous incentives for online service providers to remove content or links to content that would not be defamatory if it were assessed by a court**.

The previous tranche of changes recommended by the Model Defamation Law Working Party include important new defences, higher thresholds, and other limits to better protect freedom of expression. In order for these changes to be effective, Australian defamation law must deal with the practical, day-to-day interpretation and enforcement of defamation law by internet intermediaries. Our basic starting point is that **intermediaries must be able to deal with incoming complaints in a way that does not require them to undertake an impossible assessment of the truth of a statement or the potential applicability of other defences**.

Any serious uncertainty about an intermediary's liability will create an incentive for private actors to err on the side of caution and remove legitimate speech (for example, critical but genuine reviews). In order to enhance certainty and safeguard freedom of expression, while still providing practical remedies for people who have been wronged, we suggest the following changes:

1. Introduce **blanket immunity** for all internet intermediaries where defamatory material is posted by third parties, to avoid incentivising the removal of legitimate content;

2. Create a new **no-fault takedown scheme** for defamatory material hosted online that explicitly empowers competent courts to require internet intermediaries to remove content or links to content found to be defamatory;

3. Introduce a **new labelling scheme** for false defamatory material hosted online that requires, where practical, intermediaries to inform people accessing the material of the applicant's assertion that the truth is contested;

4. Ensure that plaintiffs can **identify prospective defendants by obtaining preliminary discovery** from internet intermediaries where there is a strong prima facie case for defamation and the court is satisfied that the orders are consistent with human rights.

Blanket immunity

Generally speaking, internet intermediaries should not be held liable for defamatory material posted by third parties. The mere fact that internet intermediaries may be able to assist plaintiffs is not sufficient to make them liable for failing to do so.[1] The core practical issue is that internet intermediaries are not in a position to assess the substance of defamation claims.[2] This is true whether intermediaries are operating technical infrastructure or operating social services. Current law creates a strong incentive for intermediaries to remove legitimate content where there is any uncertainty about the truth of the imputations or the applicability of other defences. Takedown schemes are frequently misused to convince internet intermediaries of all types to remove critical commentary or negative reviews and to silence whistleblowing or dissenting opinions.[3]

It is not possible to design a defamation scheme that imposes liability on intermediaries without creating an incentive to remove legitimate speech. In order to adequately protect freedom of expression, intermediaries should not be liable for content over which they have exercised no editorial control. Notification of mere allegations of defamation should not be enough to create liability for an intermediary with no editorial control (as is the case under our current law).

No-fault takedown scheme

We strongly recommend separating intermediary liability from the need to provide effective remedies for plaintiffs. A no-fault takedown scheme would allow plaintiffs to have defamatory content removed, without creating an incentive for intermediaries to remove legitimate content. Any legitimate takedown scheme would, however, have to be overseen by a competent judicial authority that is able to investigate questions of fact.[4] Because notice-based schemes are subject to frequent abuse, internet intermediaries should only be required to remove content upon receipt of a court order, either interlocutory or final.[5] Ensuring that courts can make enforceable orders against intermediaries who are not parties to an action would resolve the problem that would be caused by excluding intermediaries from defamation liability. Failure to comply with a takedown notice should not result in liability for defamation, but should attract a civil penalty or a remedy for contempt of court.

---

[1] Kylie Pappalardo and Nicolas Suzor. 2018. “The liability of Australian online intermediaries.” *Sydney Law Review* 40 (4): pp. 469–498. Available at: http://classic.austlii.edu.au/au/journals/SydLawRw/2018/19.html.

[2] Kylie Pappalardo and Nicolas Suzor. “Dow Jones and Company v Gutnick (2002).” In David Rolph (ed), *Landmark Cases in Defamation Law* (Hart Publishing, 2019) pp. 217-241. Available at: https://eprints.qut.edu.au/132226/.

[3] Nicolas P Suzor, *Lawless: The Secret Rules That Govern Our Digital Lives* (Cambridge University Press, 2019) (‘*Lawless*’). Available at: https://osf.io/preprints/socarxiv/ack26/.

[4] In the copyright context, see Nicolas Suzor and Brian Fitzgerald, ‘The Legitimacy of Graduated Response Schemes in Copyright Law’ (2011) 34(1) *University of New South Wales Law Journal* 1. Available at: https://www.unswlawjournal.unsw.edu.au/wp-content/uploads/2017/09/34-1-18.pdf.

[5] ‘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/.

Labelling scheme for disputed content

Removing disputed claims from the internet without a court order will always be fraught with potential risks to freedom of expression. As a legal remedy, takedown should be restricted to serious cases that have been examined in court.

Instead of a more generally applicable takedown scheme, we suggest introducing a labelling scheme for disputed content. Some classes of internet intermediaries could be required, upon request by a person who is the subject of allegedly defamatory content, to include a note alongside the content, stating that the person disputes its truth. This should only apply to intermediaries with the ability to exercise control over content; infrastructure providers, for example, should not be required to edit content that is controlled by one of their clients. Failure to comply with a request should not result in liability for defamation, but should attract a civil penalty.

Labelling schemes are a novel response to defamation, but would be consistent with evolving industry practices. For example, many major social networks have experimented with adding information labels to content about COVID-19 or contested factual assertions in news content. In the context of a highly fact-specific claim like defamation, this type of notification to users is much less likely to create chilling effects on legitimate speech.

The harm that defamation law seeks to redress is damage to reputation. For serious cases of defamation, injured people will have a remedy against the person responsible for the defamatory content, or be able to secure injunctive relief against intermediaries. For cases that are not sufficiently serious to proceed to court, it is appropriate for the range of remedies to be more constrained. In these less serious cases, a labelling scheme strikes a more appropriate balance for freedom of expression.

Identifying publishers

The threshold for making an order that an intermediary reveal a person's identity should be raised to align with the threshold applied in the UK: the plaintiff should be required to show that the prospective defendant 'arguably' defamed them, and the making of the order must be a 'necessity'.

Courts should also be required to take into account human rights and privacy considerations when deciding whether to make such an order, to ensure the orders are not abused to harass or threaten whistleblowers or other legitimate critics.