Meta response to draft Model Defamation Amendment Provisions

SEPTEMBER 2022

Meta

Executive summary

Meta welcomes the opportunity to provide comment in response to the Model Defamation Amendment Provisions released by the Meeting of Attorneys-General on 12 August 2022.

We have welcomed the changes effected by Attorneys-General in Stage 1 of the national defamation reforms and have been actively engaged in contributing to matters considered under Stage 2.

We strongly support the principles articulated by Attorneys-General in support of national defamation reform. Australia's defamation laws are not fit-for-purpose for the digital age. We agree that there should be reform to ensure "the model laws strike a better balance between protecting reputations and not unreasonably limiting freedom of expression". The seven recommendations made by Attorneys-General are thoughtful, welcome, and cognisant of concerns that Australian defamation law risks incentivising over-blocking or chilling of free speech in Australia.

Our position has long been that we support reform of Australia's defamation laws in order to clarify and expand internet intermediaries' defences around defamation content, but also to make that liability protection conditional on companies' ability to meet best practices to combat the spread of content that is defamatory.

Our previous submission provided many examples where the current regime (given recent case law) could require internet intermediaries to restrict access to online content that may benefit from being in the public domain, such as material that can generate debate that is in the public interest, but which also may be potentially defamatory, such as content relating to: the #MeToo movement; the injustices faced by people of colour highlighted by the Black Lives Matter movement or the campaign around Aboriginal Deaths in Custody; or whistleblower claims (such as negligence in the medical industry). Recent cases have created significant uncertainty for internet intermediaries by essentially requiring them to block access to content, solely on the basis of a user allegation that the content is defamatory. Notwithstanding the High Court's decision in *Defteros*, there remains uncertainty for internet intermediaries who are not search engines or otherwise make content available in ways beyond distributing a simple

hyperlink. The full rationale behind why we support reform of Australia's defamation laws can be viewed in our previous submission to the Attorneys-General's reform process.[1]

Attorneys-General have proposed a new defence for internet intermediaries (Recommendation 3), and are seeking feedback on two different approaches (referred to as Model A and Model B). Meta has previously supported the view that strengthening the innocent dissemination defence to clarify its application to internet intermediaries[2] would address these concerns (most similar to Model B, as described by the background paper). We have also provided some comments about how a safe harbour, inspired by section 5 of the Defamation Act in the United Kingdom (Model A, as described by the background paper), could be designed to be practical.

Attorneys-General have taken significant efforts in order to incorporate feedback on a safe harbour that is significantly streamlined to address concerns about the practicality of this regime. Some of these changes include: incorporating a safeguard for good behaviour by internet intermediaries; including a complete defence where the complainant can identify the originator by other means, to reduce the risks of vexatious complaints; and requiring steps to be taken in regard to a complaints notice.

Either option (Model A or B) would be an improvement on the existing state of play, given the uncertainty of current defences for internet intermediaries as evidenced in recent cases.. Given the efforts that Attorneys-General have taken to improve on a safe harbour defence from the regime in the UK, we believe Model A as described by the model provisions and the background paper would be largely workable, subject to amendments that would make a safe harbour operate properly in line with the intent expressed in the background paper.

In particular, the drafting of the safe harbour does not extend to all instances specified in the background paper. For example, it does not make the safe harbour available if the complainant does not lodge a complaints notice with the internet intermediary (ie. only sends a concerns notice or heads straight to the courts). In order for the safe harbour to operate as intended by the background paper, we recommend clarifying that the safe harbour is available if the plaintiff has not given a complaints notice to the internet intermediary.

---

[1] Meta (Facebook), *Facebook submission to the review of model defamation provisions (stage 2)*, 31 July 2021, https://www.justice.nsw.gov.au/justicepolicy/Documents/review-model-defamation-provisions/stage-2/facebook-submission.PDF.pdf.

[2] Facebook, *Facebook's response to the review of Model Defamation Provisions*, 31 January 2020, https://www.justice.nsw.gov.au/justicepolicy/Documents/review-model-defamation-provisions-amendments/facebook-submission-to-mdaps.pdf.

The drafting also involves some impractical complexity that could inhibit intermediaries' ability to respond to complaints quickly. These include:

- The different operation of the complaints notice versus the concerns notice could cause complexity and delay in responding to complaints - especially if both notices are able to be issued simultaneously. The legislation should be amended to clarify the operation between the two notices. This could be achieved via a number of means, such as by making complaints notices serve as concerns notices for the purpose of internet intermediaries, or by requiring that they are clearly sequential (with the complaints notice coming first).

- The legislation obliges internet intermediaries to maintain an easily accessible complaints mechanism. However, section 44 enables complainants to still issue a complaints notice using a wide variety of means, including via fax. In order to ensure there is no ambiguity in the process for making a complaint, the legislation should make clear that complainants are expected to use the mechanism offered by the internet intermediary. This should not be a concern for complainants, given the mechanism must be easily accessible in order for internet intermediaries to obtain the defence.

- The new provision that can hold an internet intermediary liable if they demonstrate 'malice' in relation to a complaint could be ambiguous for all parties involved. To provide greater certainty and reduce the risk of misinterpretation, the legislative provision should incorporate the guidance provided in the background paper that malice refers to the broader systems and policies of an internet intermediary, and not to any specific conduct or omission .

We recognise Attorneys-General are committed to the creation of new court powers for non-party orders for internet intermediaries to remove online content. The potential benefits of this new power, if well designed, could provide clarity for all parties by specifying and splitting the obligations of an originator of an unlawfully defamatory post compared to an internet intermediary. The non-party orders would also bring clarity and certainty that the at-issue content has been duly assessed as unlawfully defamatory under Australian law, rather than compelling internet intermediaries to make judgements based on incomplete information or lack of context.

However, as with any takedown scheme for online content, there should be strict guardrails around how non-party orders can be used. We provide some recommendations

in this submission to ensure internet intermediaries are practically able to take action in response to a non-party order, and to reduce the risk of orders that are disproportionately broad.

Finally, we would support a review of the legislation (perhaps two years after passage) to ensure it has not had unintended consequences in terms of public interest expression.

We stand ready to assist Attorneys-General in any way possible to support the rapid agreement to a national set of model defamation provisions that could improve Australia's defamation laws for the digital age.

Table of contents

Introduction

The need for reform

Meta strongly supports the principles articulated by Attorneys-General in support of national defamation reform. Australia's defamation laws are not fit-for-purpose in the digital world. We agree that there should be reform to ensure "the model laws strike a better balance between protecting reputations and not unreasonably limiting freedom of expression".

Our position has long been that we support reform of Australia's defamation laws in order to clarify and expand internet intermediaries' defences around defamation content, but also to make that liability protection conditional on companies' ability to meet best practices to combat the spread of that content. While Meta does not have editorial control over the content distributed on our services, we invest very significantly in policies and processes to combat the spread of harmful content.

Our previous submission provided many examples where the current regime (given recent legal decisions) could require internet intermediaries to restrict access to online content that is beneficial to the public, such as material that can generate public debate but also be potentially defamatory, such as content relating to: the #MeToo movement; the injustices faced by people of colour highlighted by the Black Lives Matter movement or the campaign around Aboriginal Deaths in Custody; or whistleblower claims (such as negligence in the medical industry).

In our experience, many of the individuals who make such complaints have claimed (rightly or wrongly) that this material is defamatory and should be indiscriminately removed. It is incredibly challenging to adjudicate these claims. If defamation legislation does not provide sufficient clarity for digital platforms on how to consider claims in instances such as this, it would inadvertently encourage digital platforms to excessively censor or restrict material in order to manage potential liability under the legislation.

The full rationale behind why we support reform of Australia's defamation laws can be viewed in our previous submission to the Attorneys-General's reform process.[3]

[3] Meta (Facebook), *Facebook submission to the review of model defamation provisions (stage 2)*, 31 July 2021, https://www.justice.nsw.gov.au/justicepolicy/Documents/review-model-defamation-provisions/stage-2/facebook-submission.PDF.pdf.

Meta's position in previous submissions

We agree with Attorneys-General's thoughtful and welcome discussion across multiple discussion papers about the need to clarify and strengthen defences for intermediary liability around defamatory content.

We strongly support the background paper's position that internet intermediaries should not automatically be held liable for the contents of material that is authored or created by a third party and shared on their platform. Recent cases have created significant uncertainty for internet intermediaries by essentially requiring them to block access to content, solely on the basis of a user allegation that the content is defamatory. Notwithstanding the High Court's decision in *Defteros*, there remains uncertainty for internet intermediaries who are not search engines or otherwise make content available in ways beyond distributing a simple hyperlink.

A standalone defence for digital intermediaries would reduce the risk of over-blocking and better protect free expression. It also ensures the author or creator of the material - the party with the greatest level of control over the material's contents - rightly continues to bear primary responsibility and therefore liability for defamation. The author or creator of the material is also best able to raise current defences if they choose to do so.

However, we also recognise the concerns raised by Australian policy stakeholders that defamation law should enable efficient resolution for individuals who are the subject of prima facie defamatory content on intermediary platforms .

If internet intermediaries' protection against liability for unlawful defamatory content is strengthened and clarified, we recognise that policymakers and the community may have greater confidence in the responsiveness of digital platforms by simultaneously making that liability protection conditional on companies' ability to meet best practices to both combat the spread of that content, and to swiftly restrict access to it. Instead of granting internet intermediaries an automatic, blanket immunity, akin to that which operates in the United States, an immunity for internet intermediaries in Australia could be contingent on the intermediary having adequate systems in place, such as participation in an appropriate complaints process (connecting complainants with primary publishers) or systems to restrict a piece of content in Australia, after a sufficiently rigorous independent process to determine that the content is defamatory under Australian law.

In our last submission, we took the view that the best way to achieve these objectives was to undertake reform via a hybrid model that both clarifies the application of the innocent dissemination defence to intermediaries (including by introducing a

rebuttable presumption that internet intermediaries are subordinate distributors) and by introducing a safe harbour provision that is inspired by section 5 of the United Kingdom's Defamation Act but is substantially streamlined to address concerns about the practicality of this regime.

In particular, we indicated we would only support a safe harbour provision in Australian defamation law if:

- It required the complainant to take "reasonable steps" to identify the originator.

- By taking reasonable steps, the complainant can identify the originator of the content-at-issue, the internet intermediary retains immunity until the complainant obtains judgment against the originator.

- An internet intermediary in receipt of a valid complaints notice is given a reasonable time to forward the complainant's notice to the originator (if they are able to contact them) and ask for the originator's address for service and permission to share it with the complainant. In instances where the intermediary doesn't have information that would enable them to contact the originator, the intermediary could be required to remove the content in Australia within a reasonable time, if they would like to retain the immunity.

Comments on the draft Model Defamation Amendment Provisions

Introductory comments

The provisions contains seven recommendations. We have no comments to add beyond our last submission in relation to five of these recommendations:

- Recommendation 1: Exemptions for conduits, caching, and storage services.
- Recommendation 2: Exemptions for search engine functions
- Recommendation 4: Providing an exemption from the *Online Safety Act 2021* immunity
- Recommendation 6: Courts to consider balancing factors in making preliminary discovery orders
- Recommendation 7: Mandatory requirements for an offer to make amends to be updated for online publications.

Recommendation 3: Options for a new defence for internet intermediaries

The background paper seeks comment between two alternate approaches for a new defence for internet intermediaries:

Model A: A safe harbour defence for internet intermediaries focussed on connecting the complainant with the originator. If this is not possible, the intermediary must remove the content within 14 days.
Model B: A innocent dissemination defence recognising that internet intermediaries should have a defence in defamation in relation to third-party content until the point where they are given a written complaints notice and, after that, if they remove the content within 14 days.

Attorneys-General have taken significant efforts in order to incorporate feedback on a safe harbour that is significantly streamlined to address concerns about the practicality of this regime. Some of these changes include: incorporating a safeguard for good behaviour by internet intermediaries; including a complete defence where the complainant can identify the originator by other means, to reduce the risks of vexatious complaints; and requiring steps to be taken in regard to a complaints notice.

Either option (Model A or B) would be an improvement on the current legislative framework.

Given the efforts that Attorneys-General have taken to improve on a safe harbour defence from the regime in the UK, we believe Model A as described by the model provisions and the background paper would be largely workable, subject to a number of additional amendments to improve the operation of the scheme. Model A also has the benefit of catering for scenarios where the originator of the content is identifiable and holds them responsible for their own speech.

In particular, the drafting of the defence does not extend to a full safe harbour. For example, the background paper states that the drafting provides an "automatic defence where it was possible for the complainant to identify the poster".[4] However, it is not clear whether this is presumed unless the complainant gives the digital intermediary a complaints notice.

The defence should operate as a true safe harbour, and apply at all times (provided the minimum thresholds are met). As currently drafted, this interpretation turns on the inclusion of the word "if" at the beginning of s 31A(1)(c) and should be clarified. In particular, we recommend clarifying that the safe harbour is available if the plaintiff has not given a complaints notice to the internet intermediary.

The drafting also involves some impractical complexity that could inhibit intermediaries' ability to respond to complaints quickly. These include:

- The different operation of the complaints notice versus the concerns notice is unnecessary and duplicative. If the two options were to be concurrently available, and mutually exclusive, they would create two separate regimes for notice and takedown:
    - In the case of 3A, notice via a complaint and takedown within 14 days.
    - In the case of a concerns notice, notice via the concerns notice and an offer to remove and payment of reasonable costs via an Offer of Amends within 28 days.

  This confusion is likely to slow internet intermediaries' ability to respond to complaints notice expeditiously.

[4] (Background Paper, pg. 37).

The legislation should be amended to clarify the operation between the two notices. This could be achieved via a number of means, such as by (i) whether making the complaint functions operate as a concerns notice for online publications (in which case the content of the complaint and timeframe for removal would need to be considered); or (ii) requiring the complaints notice process must occur prior to the issuance of any concerns notice and as an essential precondition to commencing proceedings.

- The legislation obliges internet intermediaries to maintain an easily accessible complaints mechanism. However, section 44 enables complainants to still issue a complaints notice using a wide variety of means, including via fax. A complaints mechanism will be more timely and effective if complainants are also required to use this process.

- The drafting grants the internet intermediary a defence if the complainant has "sufficient information to…enable defamation proceedings to be commenced". Presumably, this means sufficient information to file an originating process and *serve* an originating process. If it is intended to capture service (which we submit it should), the provision should also clarify that substituted service, including by way of email, is sufficient.

Malice

As currently drafted, s 31A(5) states that the digital intermediary defence would be "defeated" if the digital intermediary was actuated by malice[5] in providing the online service. The onus is on the plaintiff to prove malice.

The language "defeated" suggests an internet intermediary would first need to prove the defence (ie. by proving compliance with the prescribed steps following receipt of a complaints notice).

The Background Paper example suggests that, for a social media company, the relevant improper purpose must relate to provision of the service as a whole (e.g. on a policy level) rather than in relation to any acts or omissions in relation to the at-issue content. To provide greater certainty and reduce the risk of misinterpretation, the legislative

[5] According to the Background Paper, "malice" is intended to capture circumstances where the digital intermediary invited the publication with an improper motive, or created/provided/administered the forum/platform on which the matter was published with an improper motive. The Background Paper gives the example of a social media platform which encourages anonymity, etc. and has a policy not to respond to defamation reports.

provision should incorporate the guidance provided in the background paper that malice refers to the broader systems and policies of an internet intermediary.

Recommendation 5: Creation of a new power to make non-party orders for removal of content

Our previous submission outlined some reservations about the creation of a new power to make non-party orders for removal of content. However, we recognise Attorneys-General are committed to the creation of new court powers for non-party orders for internet intermediaries to remove online content.

The potential benefits of this new power, if well designed, could improve clarity for all parties by specifying and splitting the obligations of an originator of an unlawfully defamatory post compared to an internet intermediary. One benefit is that it clearly focuses on instances where the sole or primary action should be taken against the originator.

However, as with any takedown scheme for online content, there should be strict guardrails around how non-party orders can be made. Such a provision should ensure that a digital platform is in a position to actually interpret and comply with orders made from a practical perspective. Such orders should also not be overly broad. There are four concerns we have in particular:

1. We would be very concerned at any suggestion that courts should be enabled to compel blocking of material elsewhere in the world, outside of Australia's jurisdiction. Australian legal requirements for empowering courts to order removal of online material within Australia (for example, injunctions) are well established and able to achieve the public policy goals articulated in the background paper.

   To order blocking in other jurisdictions around the world would be to apply Australian legal standards outside Australia without justification, including without consideration of other legal norms that operate outside Australia. Even if courts are granted this power, other jurisdictions may (rightly) not recognise it. For example, a court in the United States would be unlikely to give effect to that order without being satisfied that the first amendment was adequately protected.

2. This new power should be *specific*, ie. only able to be applied to content that is specifically identified online, expressly by way of URL. The drafting of the new power could conceivably enable the making of orders about general classes of content or apply a descriptor to content in such a way that a digital intermediary is simply unable to interpret and comply with the order. If this occurred, it would be virtually impossible for internet intermediaries to comply with; there can be endless permutations of defamatory claims, all with different wording, and that may require specific knowledge. If the making of non-party orders risks turning into a proactive monitoring obligation for the internet intermediary, there would not only be technical barriers to our compliance with the orders, but significantly greater concerns about the risk of over-blocking and the impact on free expression.

   Applying this power to content that is specifically identified would be consistent with the complaints process in both the Model A and Model B proposals as well.

3. There should be conditions indicating when non-parties may be able to dispute the making of an order. For example, when removing the content could involve breaching a foreign law, or if it is not technically possible to restrict access in the way sought by the courts.

4. There should also be conditions indicating that a plaintiff can only seek an order against a digital intermediary where the defendant is not willing or able to remove the content.