Reset.
AUSTRALIA

Response to the Legal & Constitutional Affairs Legislation Committee Inquiry into the Social Media (Anti-Trolling) Bill, 2022

Reset Australia
February 2022

Executive summary

The recent High Court case Fairfax Media Publications vs Voller has created considerable uncertainty for Australian social media users and we welcome a legislative response to this.

The Social Media (Anti-Trolling) Bill in February 2022 as drafted provides some useful responses to this, but misses the opportunity to ensure that platforms – with their considerable power and resources – are required to undertake due diligence to reduce the amount of defamatory content promoted and shared on their platforms. Instead, this Bill focuses only on creating individual legal remedies (which may be inaccessible to many people), on an individual 'piece of defamatory content' by 'piece of defamatory content' basis.

Reset recommends:

- Only reducing liability on platforms where they have taken fair and reasonable steps to minimise the risks of defamation on their platforms, as well as implementing a complaints system.
- Requiring transparency about these reasonable steps and the complaints system through an annual reporting process
- Empowering and resourcing an existing regulator to ensure adequate oversight of these reasonable steps and the complaints scheme
- Ensuring the complaints scheme is age appropriate, and offers particular protections to those under 18 who may post defamatory content.
- Revising the scope of the Bill so it offers protections to Australians using digital products with similar functionalities to social media platforms

Contents

1. About Reset Australia & this submission

Reset Australia is an independent, non-partisan policy think tank committed to driving public policy advocacy, research, and civic engagement to strengthen our democracy within the context of technology. We are the Australian affiliate of Reset, a global initiative working to counter digital threats to democracy.

This submission has been prepared in response to the Legal & Constitutional Affairs Legislation Committee Inquiry into the Social Media (Anti-Trolling) Bill in February 2022. It documents what Reset Australia believes is the need for an overarching tech regulatory framework in Australia, and the role this Bill may or may not play in this.

2. Reset's response to the Bill

Reset Australia welcomes the Australian government commitment to keeping Australians safe from online harm, including the harmful impacts of defamatory material posted anonymously on social media. The recent High Court case Fairfax Media Publications vs Voller has created considerable uncertainty for Australian social media users and, like the government, we do not think it is appropriate that individuals and small business owners face potential defamation liability for material posted on their social media pages by others.

This proposed Bill however, must sit alongside existing and emerging regulation as part of a regulatory framework that keeps Australian's safe online and helps to realise their rights. This framework is in some places well defined, and in others, offers patchy protections.

Reset Australia has previously outlined five directions of for future policy to ensure Australia arrives at an effective, coherent tech regulation framework:

1. Enhancing the focus on systems and processes: Regulation is more effective where it focuses on eliminating risks from systems and processes, expanding on our current focus on content moderation
2. Expanding the nature of the harms address, to include community and societal risk: Regulations are currently focussed on individual level harms, but need to expand to also addresses community & societal risks
3. Requiring platform accountability and transparency: Regulation will be more effective where it requires accountability & transparency from digital service providers, rather than placing burdens on individuals
4. Requiring comprehensive coverage: Regulation should aim for comprehensive coverage, and address existing gaps and disjunctures
5. Enforcing regulations with strong regulators: To be effective, regulation needs to strong, moving beyond from self- and co-regulation, and overseen by well resourced and joined up regulators

Appendix A provides more details.

These principles create five evaluative criteria for regulatory proposals from Reset's perspective.

1. Does the Bill Focus on Systems and Processes?

- The Bill aims to address the systemic problems caused by the lack of platform liability. However, instead of creating liabilities or responsibilities on platforms it places the onus on individual posters of defamatory content. This downplays the significant role that platforms themselves play in encouraging and amplifying defamatory discourse. Platforms manufacture demand for and amplify defamatory content. They hand trolls and other bad actors the tools they need to cause harm, and directly provide incentives, including funding[1], to encourage their ongoing poor behaviour. Regulation could instead pivot to, for example, requiring platforms to change their content recommender algorithms to demote defamatory speech where it arises or reviewing their ad revenue system to ensure they do not deliberately monetise and fund defamation.

- The Bill will also operate on an 'individual piece of defamatory content' by 'individual piece of defamatory content' basis. It will leave Australians playing whack-a-mole with libellous content. A more upstream approach could be found.

2. Does the Bill address societal and community harms?

How do societal and community harms differ from individual harms?

Specific communities, such as indigenous communities, migrant communities, women, children and LGBTIQ+ people often suffer unique and disproportionate harms on social media. Disinformation and hate speech can affect communities in ways that differ from individuals.

Society itself can also be harmed by social media. The scale and reach of platforms means they can influence institutions, such as Parliament, the press and healthcare systems, often in destabilising ways. For example, social media platforms have been used to undermine public health messaging around vaccine roll out, and foreign bots have engaged in Australian electoral discussions[2], undermining our democracy.

- This Bill replicates the individual focus apparent within Australia's existing digital regulations, and does not address societal and community risks. This Bill enables individual remedies to be sought through civil cases around libel. Where the harm is broader, and society or community harm occurs, it will not provide remedy.

[1] Karen Hao 2021 'How Facebook and Google Fund Global Misinformation' *MIT Technology Review* www.technologyreview.com/2021/11/20/1039076/facebook-google-disinformation-clickbait/
[2] Select Committee on Foreign Interference Through Social Media, Senate, 30 July 2021

- Specific caution would need to be given when children are the poster of defamatory material. Firstly, children should never be asked to consent to hand over their contact details to strangers in the functioning of a platform's complaint scheme. Secondly, a higher threshold for caution must be given in issuing an EIDO against a child. We note that the recent draft Enhancing Online Privacy Bill suggested using children's 'best interests' as an appropriate test to gauge whether data processing is fair and reasonable. It would be inconsistent to realise children's best interests in the Enhancing Online Privacy Bill but not here. We would suggest the same approach in deciding whether or not a court should disclose a child's personal identification data.

- The way that Bill has been portrayed often presents an 'in principle' threat to anonymity on the internet. Anonymity is important for many different communities and user groups, including for example women, LGBTIQ+ and communities of colour. While this Bill may not meaningfully undermine anonymity 'in practice', it is important to note that anonymity is a vital element of online safety for many different communities.

3. Does the Bill increase platform accountability and transparency?

- Where a platform fails to implement an adequate complaints scheme, it will be held liable for publishing defamatory content. This creates an obligation on platforms, and some accountability, but in most instances platforms will be able to avoid liability to individual users. This limits the effectiveness of the legislation. Instead, platforms should be shielded from liability as publishers only where they have taken fair and reasonable steps to minimise the risks of defamation on their platforms (i.e. a focus on reducing risks in systems and process).

- It is unclear if platforms will be required to share data, or release periodic or other reports, about the number of requests made or their outcomes. Similar requirements exist under the public complaints facility made possible by the *Online Safety Act*. More transparency requirements could be included in the Bill.

- It is unclear if there will be regulatory oversight of how platforms operate their complaints systems, so that they can be held to account if these fail. Regulatory oversight should be a part of the process.

4. Will the Bill provide comprehensive coverage for Australians?

- The Bill largely applies to social media services, as defined under the *Online Safety Act* 2021. This means Australians will not be considered the publishers of material posted by third parties on these platforms and if they believe they have been defamed will be able to ask the platform, or the courts, to disclose the identity of who wrote the content. But Australians may enjoy these protections on similar digital services, such as online gaming networks or public messaging services. Australian's move between

digital services and platforms seamlessly, but this Bill would fail to offer protections that travel with them. If passed, this Bill should consider a broader scope potentially aligned with the definition of Social Media and Relevant Electronic Services as described under the *Online Safety Act*.

- The Bill provides for two remedies to those wishing to identify people from within Australia who may have defamed them. One requires those who post the alleged defamatory content to consent to sharing their contact details, and the other requires a court order. It is unclear how these mechanisms improve existing options.
    - If the poster of a piece of defamatory content is willing to share their contact details, they are already able to share their contact details. It is unclear how a voluntary request from a platform will create a significant change, more than a voluntary request sent to the poster directly.
    - Where court interventions are required, this duplicates existing powers. No one is truly anonymous on social media platforms, who vociferously hoover up personal data from GPS locations to telephone unique ID codes multiple times per minute, and social media companies already have to obey all lawful requests from courts or law enforcement to share data.

  This Bill purports to 'unmask trolls' but it does not substantively add to existing mechanisms.

- It is also worth noting that this Bill provides avenues for people who believe they have been defamed to take individual legal action. This approach is extremely time-consuming and expensive. It is unlikely that this will afford Australians comprehensive protections, given a significant proportion of the population is unlikely to have the resources to pursue individual legal action.

5. Will the Bill lead to strong and well enforced regulations?

- It is unclear if any regulator will have oversight of the complaints scheme, or if people will be able to seek redress if the complaints scheme fails to meet the prescribed requirements in practice. How complaints systems operate needs to be overseen by a regulatory body, and people must be able to lodge complaints about failures of the complaints system itself.

- Without regulatory oversight, a complaints scheme would simply be another self-regulatory initiative, replicating all of the known problems of Big Tech self-regulation that have led to the current circumstance.

3. Recommendations

Legislation is necessary to avoid the uncertainty and individual liabilities created by the Voller case. However, this draft Bill misses a number of opportunities to address the core issues driving defamation in digital environments.

There are a number of ways it could be improved:

- It is unclear what additional practical changes and powers the complaints scheme as laid out in the Bill would actually produce. Instead, platforms should be shielded from liability as publishers only where they have taken fair and reasonable steps to minimise the risks of defamation on their platforms. Such fair and reasonable steps could include measures such as reviewing content recommender algorithms to demote defamatory speech and ensuring platforms aren't providing ad revenue funding to 'click bait farms' that specialise in pedalling defamatory content. This risk mitigation approach could potentially be an additional requirement alongside the complaint scheme.

- Annual reporting requirements should be considered to ensure oversight of:
    - Details of, and risks assessment around, what fair and reasonable step have been taken by each platform to mitigate the risk posed by defamatory material on platforms
    - If the complaints scheme is implemented, details about the functioning of their complaint systems, including how many people made an initial requests, how many posters were located overseas or in Australia, how many posters consent to have their contact details shared, how many EIDOs were issued, the time frames for responses from platforms, and the number of complaints received about the complaint process itself.

- Empowering an existing regulator to ensure adequate oversight of the risk assessments, and complaints scheme if it is implemented.

- If the complaints scheme is implemented, ensuring the prescriptions for the scheme are age appropriate. Requirements ensuring that children are not asked to consent to hand over their contact details must be included. Further, a higher threshold for caution must be given in issuing an EIDO against a child. The 'best interests' principle provides a potential test to ensure children remain safe. The Children's Commissioner could be consulted to explore how the Bill may affect children and how to ensure their best interests are realised.

- The Bill should consider a broader scope potentially aligned with the definition of Social Media and Relevant Electronic Services as described under the *Online Safety Act*.

4. Appendix: Five policy directions for digital regulation in Australia

1. Eliminating risks from systems and processes

Regulation should pivot towards targeting risks created across the systems and processes developed by digital services. The aspects of systems and processes, and related risks, that regulation could address includes:

- **Algorithms.** These drive much of the content delivery in social media platforms, both in terms of content and advertising. For example, YouTube estimates that 70% of content viewed on their platform is as a result of their recommender algorithm and autoplay. These systems, designed by platforms, often using machine learning or other AI technologies, often promote risky or harmful content. Yet algorithms are not trained in ways that consider risks.

- **Platform design.** The user interface and user experiences of social media platforms are highly curated and engineered: each design element reflects a decision point made by a company. Platforms can be designed in ways that create risks. For example, many platforms design their user journey in ways that maximise data extraction from your device, social media apps and other internet based activity. For example, apps that nudge you to, or automatically connect with, your address book or track your GPS location. This data is used to preferences and interests, and personalise your ad experience (termed 'surveillance advertising'). This is a 'dark pattern' that maximises profits but does not consider the data risks it creates.

- **Specific features.** Specific features can also create risks. For example, features that enable the live broadcasting of locations, or photo filters that make people appear thinner. These features can combine in ways that amplify or create new risks. For example, video live streaming and the ability to receive messages from stranger's accounts creates unique risks for young users[3]. Features are developed and refined by platforms to meet identified priorities, such as maximising engagement, growing reach or extending the amount of time users stay on a platform. These priorities often do not consider risk; if 'minimising risk' was a systemic design aim many features would operate differently or be abandoned.

These sorts of systems and processes manufacture and amplify risks but none of them are inevitable. Social media platforms can change and improve their systems, and regulation can encourage them to do so.

Regulatory approaches that take a more narrow focus on content moderation (focusing on takedown/deletion of harmful or illegal content for example) are not systemic enough, nor are they commensurate with the scale of the problem at hand. They doom regulators to a perpetual game of content 'whack-a-mole' on an impossible scale.

[3]See for example, The Times' investigation in grooming via YouTube livestreams. Harry Shukman 2018 'Predators coax children into exposing themselves' *The Times* www.thetimes.co.uk/article/predators-coax-children-into-exposing-themselves-lfws0fjdp

2. Expand regulations to address community & societal risks

The risks addressed by existing legislation are too narrow, and this leaves Australians vulnerable to collective risks. Collective risks come in two interconnected forms.

Firstly, there are risks posed to specific communities, such as indigenous communities, migrant communities, people of colour, women, children and LGBTIQ+ people. These communities often suffer unique and disproportionate harms in the digital world. While some of the risks they face may be addressed by regulation around individual harms, an 'offensive-piece -of-content' by 'offensive-piece-of-content' approach can miss the collective nature of the problem. Disinformation and hate speech can affect particular communities in ways that differ from individual harm.

Secondly, platforms create societal risks. The scale and reach of social media platforms has the capacity to influence and affect Australian institutions, such as Parliament, the Press and healthcare systems, often with destabilising effects. For example, we have seen how social media platforms have been used to undermine public health messaging around vaccine roll out (often in ways with particular consequences for marginalised communities), and foreign bots engaged in Australian electoral discussions. This is not the stuff of 'conspiracy theories'; a 2021 Senate hearing revealed that Australia has been the target of a number of sophisticated foreign disinformation campaigns, including a network linked to marketing firms based in the UAE, Nigeria and Egypt, all enabled by platforms[4]..

Expanding the definitions of harms (and risks) addressed in Australia's regulatory framework would better protect Australian communities and society at large. This means tackling mis and disinformation, and explicitly addressing hate speech. Currently mis/disinformation is covered by a co-regulatory Code that has been widely criticised as 'not meeting expectations' including by regulators[5].

3. Ensure regulation creates accountability & transparency

There are multiple ways governments can regulate the digital world, but the most effective policies require accountability and transparency from tech platforms themselves. Regulations that identify the core risks as stemming from platforms themselves — and squarely place the burden of responsibility on digital services — should be prioritised.

Regulation can place duties on users in multiple ways, but these are often inappropriate or ineffective:

- Solutions that position individual users (especially children and parents) as key actors in the frontline of improving safety are often inappropriate and will fail to protect all Australians. The scale of the risks created by platforms exceed the capability of individuals to effectively manage in isolation, especially for children. The ability to 'change settings', 'effectively report content' or 'turn on safe search' will not be enough.

[4] Select Committee on Foreign Interference Through Social Media, Senate, 30 July 2021

[5] Zoe Samios & Lisa Visentin 2020 'ACMA: Tech giants' code to handle fake news fails to meet expectations' *SMH* www.smh.com.au/politics/federal/acma-tech-giants-code-to-handle-fake-news-fails-to-meet-expectations-20201026-p568oq.html

User's informed choice around settings and options is necessary, but it is not sufficient to ensure safety, particularly for those lacking the capabilities or support to do so

- Solutions that pass responsibility on to users (as parents or consumers) to read 'the fine print' or consent to a risky system misrepresents the power asymmetry between users and tech companies. The nature of the global digital architecture, and its utility in everyday life, means that withdrawing consent is not a viable option for most Australians. For example, 75% of the world's most popular million websites have google analytics and trackers built into them[6]. A 'buyer beware' approach will fail where users have no viable alternatives

- Solutions that position individual users (be they 'trolls' or influencers) as the key actors responsible for harm undersells the role of platforms in creating the risky digital environments that enable and encourage toxic actors. Platforms manufacture and amplify harmful content; they hand trolls and other bad actors the tools they need to cause harm and provide incentives, including funding[7], to encourage their ongoing poor behaviour

Accountability means that platforms themselves should have responsibilities to mitigate risks, and should be held to account where they fail and harm occurs.

Accountability also requires transparency. Part of the problem of making social media safe is that legislators, regulators, researchers and civil society often do not know enough about the specific mechanics of how platforms work nor their consequences. Requiring transparency through, for example, algorithmic audits and impact statements could help remedy this.

4. Ensure the regulatory framework is comprehensive

The rapid growth of the technology has seen Australia's issue-by-issue (e.g. 'cyber bullying', 'trolling' etc), sector-by-sector (e.g. 'social media platforms' 'messaging services' etc) regulatory framework struggle to keep pace. Many new and emergent technologies are missed, and innovative companies straddling the gaps between existing industry definitions are inappropriately regulated.

A. Gaps between industries and services

A sector-by-sector approach can fail to adequately address the shared functionalities and integration between the social media sector and multiple other industries. The most obvious of these issues is the integration of traditional media and social media platforms, but equally complicated functionalities exist between social media platforms and data brokers, other online services, the advertising sector, the broader telecommunications industry, and increasingly emergency services and health and social care services as they become central to public messaging campaigns (among others).

---

[6] Steven Englehardt & Arvind Narayanan 2016 'Online Tracking: A 1-million-site Measurement & Analysis' *CCS '16: Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security* doi.org/10.1145/2976749.2978313

[7] Karen Hao 2021 'How Facebook and Google Fund Global Misinformation' *MIT Technology Review* www.technologyreview.com/2021/11/20/1039076/facebook-google-disinformation-clickbait/

Current legislation oftens fails to reflect these integrations and diverse functionalities. Using Roblox, an online kids game, as an example highlights the sorts of peculiarities this can lead to. The current definition of a 'social media' company (as laid out in the *Online Safety Act* and the proposed Enhancing Online Privacy Bill) would fail to cover Roblox. Roblox allows the creation of personal avatars; facilitates and encourages interaction and communication between users, and; allows users to create and share games for others to play. Because users do not post content *per se* — they 'post' games — they are unlikely to be considered a social media platform under the existing definition. Roblox is however covered under the *Online Safety Act* as a 'relevant electronic service' as it facilitates messaging and game play between users. But it would not be covered by the proposed Enhancing Online Privacy Act. This means that because Roblox allows users to 'create and share games' rather than 'create and share content', kids may be protected against cyberbullying with regulation, but may not be protected from exploitative data practises. This is regulatory bingo.

Scope limitations can also create regulatory discrepancies between the digital and non-digital world. For example, the ACCC's Digital Platform Inquiry explored the patchwork of regulations that applied to digital publishers compared to telecoms, radio comms and broadcast industries[8]. The ACCC found that despite providing comparable services, digital platforms often feel outside the scope of existing regulation, such as press statements of principles or legislation around advertising gambling and medicine.

Likewise, some exemptions in Australian regulation have not kept pace with the changing digital world. The use of turnover thresholds, such as exempting business above or below $3m or $10m annual turnover, are blunt and do not reflect the emerging nature of the tech industry. In particular, small tech startups can create significant risks for Australians but can often be overlooked. For example, the *Privacy Act* places obligations on businesses with an annual turnover of over $3m but exempts those under this threshold — even those handling significant amounts of data. Blunt exemptions often miss risks, and can leave large companies with safe practice facing a high regulatory burden, while extremely risky small companies continue to deliver harmful products and services.

B. Gaps in emergent technologies

Likewise, an issue-by-issue, sector-by-sector approach cannot anticipate risks created by innovations and emergent technologies. This has left recent innovations unregulated in Australia, including for example; surveillance advertising and ad delivery systems; AI; blockchain and its integration across systems, and neural technologies.

These gaps suggest that the current approach is unable to future-proof the regulatory framework, and that as technologies evolve, more and more gaps will emerge. Risk focused, systemic models may be more successful at future proofing themselves.

Australians use a wide range of digital services, and seamlessly move between technologies, sectors and companies of all sizes. Their safety should be ensured across their whole digital ecosystem. Gaps and exclusions within Australian regulation have often left Australians reliant on foreign legislation for protection.

[8] Table 4.1, ACCC 2019 *Digital Platforms Inquiry: Final Report* www.accc.gov.au/system/files/Digital%20Platforms%20Inquiry%20-%20Final%20report%20-%20part%201.pdf

5. Ensure regulation is strong and enforced

Big tech poses big risks and necessitates a robust regulatory response. However, because Australia has to date engaged self- and co-regulatory models by default, our regulatory framework has often failed to reduce risks as rigorously as they otherwise may have.

Future regulation needs to start from the premise that self- and co-regulation will not be sufficient for the social media sector. Reset Australia believes self- and co-regulation have a role to play in the Australian regulatory landscape at large, but that unfortunately the risks posed by the digital environment are:

- High impact, and include significant public health and community safety concerns
- Significant to the community, and the public has an appetite for the certainty of robust regulations
- Unable to be adequately dealt with by lighter touch regulations. The social media sector has demonstrated a track record of systemic compliance issues, including multiple breaches of existing legislation and a generally anaemic response to self-regulation

This warrants a pivot towards primary and subordinate legislation and regulation for the sector.

Alongside strengthening existing regulation, regulators need to be resourced and enabled to enforce this. This includes the ability to fully utilise existing regulation as well as any new legislation proposed.