



28 February 2022

Committee Secretary
Senate Legal and Constitutional Affairs Committee
PO Box 6100
Parliament House
Canberra ACT 2600

By email: legcon.sen@aph.gov.au

Dear Chair,

Thank you for the opportunity to provide a written submission and feedback on the draft *Social Media (Anti-Trolling) Bill 2022 (SMAT Bill)*.

We share the Australian Government's desire to promote online safety, and Twitter remains focused on making people feel safe, secure, and empowered to participate in the public conversation every day.

As we continue to iterate and strengthen our approach to meet evolving contours and challenges surrounding online behaviours, we're moving with urgency, purpose, and our commitments to develop and enforce a range of policy, procedural, and product changes to help people feel safe, welcome, and to control their experience on Twitter. We support smart regulation, and our focus is on working with governments to ensure that regulation of the digital industry is practical, effective, inclusive, and feasible to implement to ensure that certain core democratic values are intact while promoting tech innovation, including Twitter's core commitment to an Open Internet worldwide.

Our submission stands together with the respective submissions from the Digital Industry Group Inc. (DIGI) and Business Council of Australia (BCA), both of which Twitter is a member. For clarity, and to complement and reinforce these statements, we've structured this submission to address the key issues within the SMAT Bill as they pertain to Twitter operating in Australia.

Twitter is committed to working with the Australian Government, our industry partners, non-government organisations, academics, and wider civil society as we continue to build our shared understanding of the issues and find optimal ways to approach these together.

We trust this written submission will be a useful input to the Committee's consultation process. Working with the broader community we will continue to test, learn, and improve quickly so that our platform remains open, accessible, effective, and safe for everyone.

Thank you again for the opportunity to input into this important process.

Kind regards,

Kara Hinesley
Director of Public Policy
Australia and New Zealand

Kathleen Reen
Senior Director of Public Policy
Asia Pacific



Introduction

In this submission, we explore concerns related to the *Social Media (Anti-Trolling) Bill 2022 (SMAT Bill)*.¹

Traditionally, defamation legislation has been governed by the State and Territories in Australia. The introduction of the Model Defamation Provisions (**MDPs**) in 2005 saw the harmonisation of these state-based laws. More recently, the Council of Attorneys-General Defamation Working Party undertook a review of MDPs to make these laws fit for purpose in a digital age. This review was split into two stages, and the Working Group made substantial progress in updating the MDPs under Stage 1 reforms. Stage 2 of the reformation process, which specifically considers the question of internet intermediary liability in defamation for the publication of third-party content, is currently underway.

Both directly and through our trade associations, Twitter has participated in the State and Territory-led review process over the past few years. We believe there is considerable overlap between the mechanisms proposed by the SMAT Bill, and those under consideration in the Stage 2 Review. The Stage 2 Review is at an advanced stage and well-placed to take reforms in this area forward.

We acknowledge that defamation laws need to be updated to account for the online environments; however, we do not believe the SMAT Bill meets the Government's stated goal of "combating online trolling."² Instead, the SMAT Bill appears to overlap with established preliminary discovery processes that are currently in place for defamation claims. Addressing instances of abuse online versus re-defining liability for online publications are two separate and complex issues of legal liability that should not be conflated and doing so risks significant harm to free expression online.

Additionally, it is not clear how this Bill intends to interact with the review of the MDPs and how the new disclosure order under the Bill would operate alongside court orders under the current preliminary discovery process that exists today. As currently drafted, the SMAT Bill put forward by the Commonwealth will overtake and subsume the State and Territory Attorneys-General active review of the MDPs, which brings forward questions regarding the separation of powers between the Commonwealth and the States and Territories.

Furthermore, the Bill brings forth uneven liability concerns between analog and digital mediums, as well as privacy outcomes. As written, this Bill could compel a service that doesn't otherwise collect certain personal information to be required to capture substantially more information to comply with this law, which stands alone from any other legislative regime.

The SMAT Bill is an example of the current policy environment in Australia, which contains fragmented responsibilities across multiple government departments and agencies. The past twelve months has seen considerable activity in the online regulatory space. Many of the consultation processes have worked to often overlapping and accelerated timelines, resulting in duplication and inconsistencies across old and new legislative regimes.

As the economy becomes increasingly digitised across all Australian industries, a more coordinated approach to developing the foundational policy and regulatory frameworks is key to ensuring success.

The Department of Prime Minister & Cabinet (**PMC**) announced its Deregulation Agenda in 2021 and made a \$120 million investment in the initiative.³ As part of the agenda, it released a Regulator Performance Guide that came into effect from 1 July 2021, which states that "best practice requires that regulators consider, and aim to improve on, the combined regulatory burden of governments on business and the community."

Governments need to take a coordinated approach with respect to the review of existing legislation and passage of new legislation, and these best practices are not observed in either the process or the sequencing of this

¹ *Social Media (Anti-trolling) Bill 2022* (Cth) <<https://www.legislation.gov.au/Details/C2022B00015>>.

² <https://www.smh.com.au/national/defamation-experts-reject-morrison-government-s-anti-troll-proposal-20220123-p59qg4.html>

³ <https://deregulation.pmc.gov.au/>



legislation as it preempts and overrides the substantial work and engagement that has taken place through the national Model Defamation Provision reform process. As currently drafted, the SMAT Bill puts the proverbial “cart before the horse” and negates the current work that is already underway.

We urge a whole-of-Government approach to assess these reforms in a macro context, and consideration must be given to whether the requirements under the SMAT Bill will help or hinder the Government’s Deregulation Agenda, alongside the many other unintended consequences outlined in this submission.

Changes to the Exposure Draft

The version of the SMAT Bill that was referred to the Australian Senate Legal and Constitutional Affairs Legislation Committee for inquiry and report introduced some key changes discussed below.

Change of terminology from defamatory “comments” to defamatory “material”

The SMAT Bill changes the terminology contained in the original Exposure Draft from defamatory ‘comments’ to defamatory ‘material.’ ‘Material’ is defined by reference to the *Online Safety Act 2021 (OSA)* in section 5 and includes ‘material’ in the form of text, data, sounds, visual images ‘or any other form.’ With this change, the Exposure Draft now neglects to define the term ‘comment’ at all within the text of the SMAT Bill. Additionally, the Uniform Defamation Law uses the term ‘matter’ to define communications for the purpose of law.

Given the three differing terms across these various pieces of legislation, we would ask the Commonwealth Government to harmonise these definitions and their scope to provide certainty for providers as to their obligations arising under state and federal law, especially as the Bill imposes a blanket imposition for liability for publication upon providers.

Amendments to innocent dissemination defence

In relation to the defences available under the SMAT Bill, the current thresholds place unrealistic requirements for businesses to access the defence. The Bill will not allow regulated entities to access a defence where an individual who has posted allegedly defamatory content declines to provide their contact details and a court declines to grant an order requiring disclosure of the individual’s contact details. The Bill continues to be disproportionately punitive, preventing business from accessing a defence because of a decision made by a third party that they have no control over, and where a court may have determined not to disclose these details.

Additionally, material privacy concerns still remain in relation to the SMAT Bill’s requirements that businesses collect substantially more personal information on all Australians that use their services. The definitions in the Bill specify that a business will need to be able to supply details of a person’s name, email address, phone number, and any other details determined by the legislative rules.⁴

The requirement for businesses to hold this information will effectively deny Australians the ability to maintain their privacy through anonymous or pseudonymous accounts. For businesses to ensure they can access the defences set out in the Bill, they will need to both confirm the accuracy of any details that are provided to ensure they can ‘actually disclose’ this information. The removal of anonymity will have a regulatory and social cost well beyond the problem the Government is seeking to solve, and it needs to be balanced against legitimate opportunities for people to exchange information, ideas, and express their opinions and beliefs.

While the Bill’s explanatory memorandum suggests that it does not require platforms to collect this information, in the same stroke, the Bill also effectively removes any avenue for these businesses to access a defence against liability for defamation claims without this type of information. The Bill, as drafted, is a blunt instrument that will run counter to the best interests of Australians absent the tangible improvements to online safety outcomes promised by the Government. Thus, we recommend the ‘innocent dissemination’ defence be maintained in its

⁴ *Social Media (Anti-trolling) Bill 2022 (Cth)* <<https://www.legislation.gov.au/Details/C2022B00015>>.



original form, and platforms be able to access a defence where they can show reasonable efforts to connect a complainant with the originator of the comment.

Beyond the costs of removing anonymity for all Australians online, the requirement for social media businesses to collect and hold even greater volumes of Australians' personal information runs counter to other government initiatives and general best practice privacy principles, such as data minimisation.

Nominated entity requirements

The SMAT Bill requires entities captured by the Bill to have a 'nominated entity' capable of meeting the obligations arising under or in connection with the Bill, which carries with it a threat of substantial penalties. These obligations include requiring the nominated entity to have access to "the relevant contact details...and country location data of end-users...in Australia."⁵

The Bill is already designed such that entities operating in Australia will be required to comply with obligations that may arise under or through the operation of the Bill. Additionally, Australia has long been a champion of rules-based frameworks that encourage digitalisation of trade and has argued against data localisation requirements. Unfortunately, the current drafting of the Bill runs counter to these principles and portrays Australia as supportive of regulatory regimes that increase barriers to cross-border data flows. Given that entities likely to be regulated by this legislation are headquartered in the United States, it also appears in breach of the Free Trade Agreement (FTA) between Australia and the United States.⁶ Article 10.5 of the FTA explicitly states that neither party will "require a service supplier of the other Party to establish or maintain a representative office or any form of enterprise, or to be resident, in its territory as a condition for the cross-border supply of a service." We recommend the Committee reconsider the intent behind this section in the Bill and the need for this requirement to achieve its intended objective.

We believe there is considerable overlap between the mechanisms proposed by the Bill, and those under consideration in the Stage 2 MDP Review. Accordingly, Twitter would recommend that the Government continue to work with industry through the mechanisms already in progress through the implementation of the *Online Safety Act 2021* and support the continuing work on the existing MDP reform process led by the States and Territories to achieve meaningful improvements in Australia's online safety and defamation laws, respectively.

Definitions in the Bill

In the current exposure draft, the SMAT Bill appears to be primarily concerned with defamation complaints and proceedings, and does not reference the act of 'trolling' anywhere except in its title and associated press releases.⁷

Australian courts have previously stated that 'mere vulgar abuse' is not defamatory if it fails to reach the standard of seriously harming a person's reputation or diminishing their standing in the community. With regards to "trolling,"⁸ it is generally defined as inflammatory, insincere, digressive, extraneous, or off-topic messages posted in an online community. While trolling can be harmful to its intended target in certain contexts, trolling *by itself* may not rise to a level sufficient to give its victim a cause of action in defamation.⁹

Additionally within Australia, there are effective means by which people can seek appropriate, timely remedies to harassment online via the recently implemented *Online Safety Act (OSA)*, which enables victims of online abuse to have content removed within 24 hours.¹⁰ Additionally, the OSA has armed the eSafety Commissioner with information-gathering powers so the eSafety Office can conduct investigations into harmful online behaviour and

⁵ Ibid, s 22.

⁶ <https://www.trade.gov/us-australia-free-trade-agreement>

⁷ *Bennette v Cohen* (2005) 64 NSWLR 81, 97-98 (Bryson JA).

⁸ https://en.wikipedia.org/wiki/Internet_troll

⁹ *Aldridge v Johnston* [2020] SASCFC 31 [119], citing Patrick George, *Defamation Law Australia* (LexisNexis Butterworths, 3rd ed, 2017) 218.

¹⁰ *Online Safety Act 2021* (Cth).



issue fines and notices. These powers grant the eSafety Commissioner the ability to issue end-user notices that require a person who posts cyberbullying material to remove the material, and enables the Commissioner to obtain identity information including basic subscriber information for anonymous accounts.¹¹ On its face, it would appear that the schemes under the OSA may be a more efficient and relevant pathway for Australians to effectively deal with online abuse rather than the other avenues contained in the SMAT Bill.

Additionally, the definition of “social media service” in the SMAT Bill is defined to have the same meaning as in the OSA, which would lead to the application of the SMAT Bill to essentially every website available in Australia that enables interaction between two users. This reading of the definition would extend the scope of this Bill and would not limit its application to just social media companies. Thus in practicality, the actual body and text of the SMAT Bill again does not appear to address online abuse, but instead captures unintended online services and focuses on defamation complaints and proceedings.

With the OSA’s recent commencement, it remains to be seen how these legislative schemes will fully operate in practice. However, it appears that there is significant overlap between these laws, especially with regards to the processes a social media service would need to have in place to respond to complaints of abusive material, and creates uncertainty with respect to the definitions and language adopted across federal legal regimes.

The Model: Complaint scheme and safe harbour

Within the SMAT Bill, the proposed model encompasses a safe harbour that is afforded to social media companies if they adopt a complaints scheme in line with the legislation. Under the proposed complaints scheme, social media platforms would be asked to hand over the identity of anonymous commenters to potential plaintiffs looking to sue those commenters for defamation. In exchange for following the complaints processes and revealing commenters’ identities, the platforms would have a new defence against being held liable for defamation for the material themselves. However, as currently drafted, social media platforms could only delete a comment or identify the anonymous author if the commenter themselves provides their consent. If the commenter did not consent to their identity being revealed, the complainant would need to get a Federal Court order to require the platform to disclose that information.

Twitter is concerned that the obligations imposed by these mechanisms are unclear and do not appropriately address circumstances where the requisite consent is not provided by a person that posts defamatory material, or where a person who posts defamatory material cannot be located, or is outside of Australia.

We are also concerned about the process for compliance proposed in the complaints scheme, as companies may face circumstances where the proposed obligations under the Bill conflict with other laws relating to privacy and data in other jurisdictions.

As it stands, Australian courts have a process for unmasking online accounts. Australians can apply for preliminary discovery in court, requesting that companies provide details to a claimant. Therefore, we question why a new order under the SMAT Bill is needed, and how the process in the draft Bill would make it more expeditious for people to sue for defamation in Australia beyond the legal frameworks that are already in place.

Additionally, a court order under the SMAT Bill would only be available in the Federal Court of Australia or the Federal Circuit and Family Court of Australia despite defamation cases being typically brought in the State and Territory District and Supreme Courts.¹² This new process directed orders through the federal court system could cause further costs, delays, and onerous processes for potential claimants under this Bill.

Innocent dissemination defence

¹¹ <https://www.esafety.gov.au/sites/default/files/2021-07/Online%20Safety%20Act%20-%20Fact%20sheet.pdf>

¹² *Social Media (Anti-trolling) Bill 2022* (Cth), s 26.



Section 15(3) of the Bill prevents a provider from relying on the defence of innocent dissemination in circumstances where the provider is a publisher of the comment and is a party to defamation proceedings related to the comment. Section 15 of the Bill provides a defence to a provider that complies with the section 16 complaints scheme or an end-user information disclosure order under section 19.

Our reading of the current drafting of the Bill is a comprehensive removal of the innocent dissemination defence. Despite the intentions of the Bill, it is possible for a prospective applicant to commence proceedings without first engaging in the resolution mechanisms, most notably through the issuing of a compulsory concerns notice as is required to commence proceedings under State or Territory legislation. This would be problematic for providers, as it is not mandatory for a complainant to engage a resolution mechanism under the Bill before commencing defamation proceedings against a provider, leaving a provider without access to a defence as contemplated by the Bill, nor the defence of innocent dissemination. This is an inequitable outcome for a provider who is joined to defamation proceedings in which the proper resolution mechanisms have not been engaged.

Additionally, the idea behind withdrawing the innocent dissemination defence is to ensure there is always a defendant available to make a defamation claim against, which is questionable. The defence is now available where the social media service does not have knowledge of the claim regardless of whether it is treated as a publisher or not. In the High Court's decision in *Fairfax Media Publications v Voller (Voller)*, the court also confirmed that the availability of the defence is not linked to publisher status.¹³ Therefore, the complainant end user should still be required to notify the social media service of a claim before the innocent dissemination defence is withdrawn.

It is also likely that prospective complainants may not wish to undergo either of the resolution mechanisms provided in the SMAT Bill for a range of reasons. Most notably, they may not wish to subject themselves to additional processes to find the individual commenter when a cause of action lies against a social media provider as a publisher pursuant to section 15 of the Bill. The legislative mandate that providers are publishers, and that the defence of innocent dissemination does not apply (whether or not the resolution mechanisms were engaged) is likely to contribute to a surge in litigation in Australian courts by encouraging prospective applicants to commence proceedings directly against a provider rather than attempting to identify and join the individual who posted the defamatory content. This would have the opposite effect of the Government's publicly-stated intention, which is to ensure that post-Voller, *individual commenters* (i.e. trolls) are held accountable for their comments or material posted online, and would result in a bypass mechanism rather than bringing wrongdoers to justice, which is the intended purpose of the legislation.

Clause 16: defence for the provider of a social media service

Clause 16 of the Bill creates a defence for the provider of a social media service in a defamation proceeding in which they are a defendant in the scenario where a user has made a complaint to the provider about defamatory material, the provider has complied with the complaints scheme in relation to that defamatory material, including by disclosing the country location data of the person who posted the material (i.e. the poster), and if the following conditions are satisfied:

- the user has not requested the social media services provider to disclose the relevant contact details of the poster and an end-user disclosure order by a court has not been made in relation to the defamatory material;
- the user has requested disclosure of the poster's relevant contact details under the complaints scheme, and the provider has disclosed the relevant contact details to the applicant; and
- where an end-user disclosure order has been made as a result of the defamatory material, the provider has disclosed the required information.

The steps required to be taken by a provider in this process to avail themselves of the defence in Clause 16 prove to be very problematic as they essentially confer investigative powers on end user complainants (i.e. through demanding user information of an original commenter). This would be a significant deviation from

¹³ *Fairfax Media Publications v Voller* [2021] HCA 27.



globally accepted norms surrounding disclosure of user data. Ideally, disclosure requests should be made by courts and/or authorised investigative authorities in ordinary circumstances to protect against potential misuse and preserve inherent privacy rights.

Then complex issues arise when the defence is not available because the originator of the comment has not consented to disclosure of their contact information, or where the social media provider cannot comply with a court order as it simply does not have the necessary contact details at its disposal. Section 17 of the Bill provides that upon the request by a complainant for the commenter's contact details, the social media provider may only disclose such details with the commenter's consent. A defence under section 16 of the Bill is not available to a provider where the provider has not received consent for disclosure, and the complainant has not subsequently applied for or been granted an end-user information disclosure order under section 19.

As drafted, social media services cannot get advance consent to disclose users' contact information via terms of conditions of service or a privacy policy. The requirement is for consent to be expressly sought and provided on each individual basis. Thus, the draft Bill as currently proposed does not provide adequate safeguards or protections for businesses.

It must also be noted that by granting powers to end-users to demand a poster's information means that we are delegating investigative powers of law enforcement and courts to end-users, which would mean we are diluting the supremacy of institutions such as the judiciary.

Given that disclosure of contact details is for the purpose of providing a complainant with a means for redress against the poster directly, Twitter anticipates that the poster will invariably not consent to the provider providing the contact details to the complainant. It is, therefore, our view that in these circumstances, the defence should be available to providers who have complied with all relevant requirements in seeking to obtain consent, whether or not consent is ultimately provided. This will ensure further certainty for providers that compliance at all stages will guarantee them the benefit of the safe-harbour defence in section 16.

Additionally, it is important to note that Stage 2 Review of the MDPs also contemplates the introduction of a safe harbour defence subject to a complaints notice process and considers this in light of potential reforms to the innocent dissemination defence. Instead of imposing liability for publication on providers, but not page owners, as contemplated by section 14 of the Bill, the Stage 2 Review considers how the innocent dissemination defence may be amended to take into consideration the role of page owners and providers in the publication process. There may be circumstances where the blanket removal of an innocent dissemination defence would result in unfair and unintended outcomes, and given these substantial concerns, there is merit in awaiting the Stage 2 Review's recommendations in this respect.

Clause 17 and 19: Complaints scheme, end user information disclosure orders, and prescribed requirements

The SMAT Bill requires a provider to provide the contact details of the commenter in compliance with both the complaints scheme (s 17) and end-user information disclosure orders (s 19). Under either mechanism, the contact details to be provided are the commenter's name, email address, and phone number.¹⁴ The Explanatory Paper states that the details "need to be effective to contact the commenter. Details that turn out to be fake will not allow the provider to access the defence."¹⁵ The Bill incentivises services to collect more personal information from their users by offering legal safe harbours if they do so. This means that Australians who prefer not to use their real name online, operate anonymously or pseudonymously, or choose not to share their contact details with a very wide range of websites for safety or privacy reasons, may no longer be able to do so.

Anonymity and pseudonymity on Twitter

¹⁴ Social Media (Anti-Trolling) Bill 2021, s6 definition of 'relevant contact details'.

¹⁵ Social Media (Anti-Trolling) Bill 2021: Explanatory Paper, p. 4-5.



Abuse and harassment have no place on Twitter, and we have robust policies and enforcement mechanisms to prohibit and rapidly remove harmful content from our platform under the Twitter Rules.¹⁶ As an open service, our rules reflect the voice of the people who use Twitter.

Our aim is to have comprehensive policies that appropriately balance fundamental human rights and consider the global context in which we operate. We have recognised the role of deep consultation to appropriately address the complexity of online safety issues. Through the development of our policies, products, and partnerships, we have undertaken coordinated efforts to consult with a range of partners, including human rights experts, civil society organisations, academics, the general public, and our Trust & Safety Council, whose feedback is reflected in revisions to the policy frameworks that govern the Twitter platform.¹⁷

We want the Twitter Rules and Terms of Service (TOS) to ensure all people can participate in the public conversation freely and safely. We take our role in promoting healthy conversation seriously, and as part of that mission and based on feedback from partners, we allow people to use Twitter pseudonymously or as a parody, commentary, or fan account.¹⁸

Our partners have emphasised that anonymity and pseudonymity can be a critical entry point for people to connect online and build trusted relationships. Trust can be derived from the content someone posts and how they connect with others.¹⁹ Trust is based on what you do, not who you are, and anonymity or pseudonymity provides space for more people to express themselves freely and safely, in ways that actually engender that sort of trust-building connection. Simply put, in addition to providing safety, anonymity and pseudonymity provides people with the agency and control to choose how they present themselves. This has been a core tenet of the internet since its inception and is essential to a society that promotes individual choice and freedoms.

To be clear, pseudonymity is not a shield to the Twitter Rules or TOS, nor is it a shield from criminal liability. When a person signs up for a Twitter account, we ask for a verified email or phone number, which can assist our teams in preventing ban evasion, platform manipulation, or compliance with duly executed legal and information requests.²⁰ Police and authorised representatives can submit requests for information about an account they're investigating any time via our dedicated Legal Requests Submission site.²¹

Twitter does not allow fake accounts or platform manipulation on our service. We work to prevent spam and fake accounts from harassing other people on the service both at the sign-up stage so they won't be able to join, and by removing accounts that have been proven to cause trouble. Twitter detects roughly 25 million accounts per month suspected of being automated or spam accounts.²² During the first half of 2021, we issued 130 million anti-spam challenges and observed an approximately 10% increase in the number of spam reports from the previous reporting period.²³

Notably, various academic studies have already shown that anonymity alone does not lead to harassment, and in many cases, people post harassing, toxic replies under their real name with a photo of their real face.²⁴ Research into why people might harass others on the internet, and how to most effectively tackle the behaviour, continues to evolve, including by deepening our understanding of the social contexts these problems exist within. Currently, there is not conclusive evidence that requiring the display of names and identities will reduce

¹⁶ <https://help.twitter.com/en/rules-and-policies#twitter-rules>

¹⁷ <https://about.twitter.com/en/our-priorities/healthy-conversations/trust-and-safety-council>;
https://blog.twitter.com/en_us/topics/company/2019/synthetic_manipulated_media_policy_feedback

¹⁸ <https://help.twitter.com/en/rules-and-policies/platform-manipulation>

¹⁹ <https://blog.twitter.com/common-thread/en/topics/stories/2021/whats-in-a-name-the-case-for-inclusivity-through-anonymity>

²⁰ <https://help.twitter.com/en/rules-and-policies/ban-evasion>

²¹ https://legalrequests.twitter.com/forms/landing_disclaimer

²² <https://transparency.twitter.com/en/reports/platform-manipulation.html#2020-jul-dec>

²³ <https://transparency.twitter.com/en/reports/platform-manipulation.html%232021-jan-jun&sa=D&source=docs&ust=1643346342042928&usg=AOvVaw3zYJk7mDfSTPqZkjPwjFPJ>

²⁴ <https://jilliancyork.com/2021/01/14/everything-old-is-new-part-2-why-online-anonymity-matters/>



social problems, and many studies have documented the problems it actually creates, like posing real threats to vulnerable communities.²⁵

Additionally, the potential consequences involved with digital identification policies might have on online participation, access, and the widening of the ‘digital divide’ should be considered. Personal identification can pose risks to vulnerable groups who are not able to safely use services under their real name, such as those seeking information or support for domestic violence, whistleblowers, or LGBTQIA+ people. The requirement for digital identification verification could also cause inadvertent repercussions in people accessing services on the internet. According to the World Bank, an estimated 1 billion people worldwide do not have an official form of identification.²⁶ It is often the marginalised, vulnerable, and impoverished who lack government-issued IDs, leading to larger inequalities amongst people being able to access online services.

With regards to the SMAT Bill and Stage 2 of the MDP reform process, unlike the Bill, the Stage 2 Review suggests the mechanism of court orders to identify originators of defamatory content; however, it makes reference to countervailing considerations, like privacy protections and whistle-blower protections. As the Electronic Surveillance Reform process²⁷ and the *Privacy Act* reform process²⁸ are also taking place simultaneously, which involve related privacy and surveillance concerns, we believe further consideration ought to be given to these issues, particularly in relation to sections 17 and 19 of the Bill.

Such regulatory proposals to collect additional information also run counter to the universally accepted privacy practice of data minimisation. Data minimisation requires goods or service providers to not seek to collect data beyond what is reasonably needed to provide the good or service. Data minimisation that forms part of the existing APPs under the *Privacy Act 1988* (Cth), and is also a key principle of the Consumer Data Right.²⁹

Verification of contact information for accounts

In addition to the issues arising from additional information collection, the process of verifying a commenter’s details is a complex one, as people frequently change emails, phones, etc., and the current draft Bill places an onerous obligation on providers to verify details or risk losing the benefit of the defence.

Also, where a commenter does not provide consent to disclose their contact details, and an end-user information disclosure order is made under section 19 of the Bill, it may not be a straightforward process to verify that the information the provider has on record is correct. The proposed scheme places obligations on providers that extend well beyond the contemplated scope of holding commenters of defamatory material to account, imposing an onerous and ongoing obligation on providers to constantly verify the identities of all users, and update that verification with the cooperation of users on an ongoing basis. This obligation imposes an unreasonably high and challenging technical responsibility for providers to meet.

Therefore, the SMAT Bill must contemplate inevitable situations where the contact information held and accessed by the provider will not be the correct contact information of the commenter. In these circumstances, if the provider undertakes all reasonable steps in the complaints process in compliance with the order, the defence under the Bill should be made available to the provider, including in circumstances where a commenter’s contact details cannot be verified.

Furthermore in line with the principle of proportionality, once the contact details are provided to the complainant, the onus should then shift to the complainant to make all reasonable attempts to contact the commenter before the institution of proceedings. In the absence of such an obligation on the complainant, the provider may be

²⁵ <https://coralproject.net/blog/the-real-name-fallacy/>

²⁶ <https://id4d.worldbank.org/global-dataset>

²⁷ <https://www.homeaffairs.gov.au/about-us/our-portfolios/national-security/lawful-access-telecommunications/electronic-surveillance-reform>

²⁸ <https://www.ag.gov.au/integrity/consultations/review-privacy-act-1988>

²⁹ OAIC, “Chapter 3: Privacy Safeguard 3 — Seeking to collect CDR data from CDR participants”, accessed at <https://www.oaic.gov.au/consumer-data-right/cdr-privacy-safeguard-guidelines/chapter-3-privacy-safeguard-3-seeking-to-collect-cdr-data-from-cdr-participants/>.



denied a defence despite having complied with all obligations, simply due to a complainant's limited efforts to contact a commenter have been unsuccessful.

Assessing likelihood of complainant's right to obtain relief

The obligations in the SMAT Bill on social media services apply in circumstances where the complainant end user has 'reason to believe' that they may have rights against a commenter in defamation proceedings, i.e. the complainant has a right to obtain relief. Assessing the likelihood of relief being granted to a complainant is a complex undertaking, which requires consideration of the nature, scope, and broader context of the publication in question; hence why this is usually undertaken through a court process with evidentiary standards and rules of procedure.

The inherent complexities of defamation law mean that members of the public are not necessarily in a position to determine whether a comment made online is truly defamatory and has caused them sufficiently serious harm, so as to give rise to a right to relief. As you may be aware, the law of defamation itself is an evolving concept which is backed by years of jurisprudence being laid down by courts. Social media providers now seem to be tasked with dispensing the role of judiciary as to determine what content is defamatory, which should be best left to courts.

We would query whether the Bill contains adequate safeguards to stop malicious actors attempting to game the complaints scheme to gain personal details and mis-using those details for malicious purposes. Under section 17(1)(d), the Bill would require a provider (with no contextual information about the publication or parties involved) to disclose to a complainant, within a short 72-hour timeframe, a poster's country location data and under section 17(1)(g)(i) seek consent from a poster to share their relevant contact details with a complainant. We are concerned that this process could be utilised for improper purposes (e.g. stalking, vexatious or politically motivated motives).

Section 17(1)(i) permits providers of social media services not to take action under the complaints scheme if they reasonably believe that the complaint does not genuinely relate to potential defamation proceedings against the commenter. However, it is unclear whether providers are required to accept the complainant's opinion as to whether they have a potential cause of action, or whether the provider is able to make its own assessment as to whether the complainant has a reasonable belief that they may have a cause of action in defamation, which would require additional time and resources for a thorough review of each report, and would be challenging with the tight turnaround times contained in the Bill.

Therefore, it seems victims of trolling may either be unable to rely on the mechanism outlined in the SMAT Bill's model due to either: (1) not having a right to relief in defamation; or (2) they will need to seek legal advice in order to determine if they they have a right to obtain relief, and would incur expenses in doing so (which the Bill states it's trying to avoid), as as many hurtful comments or material may fall well below the threshold for defamation.

Additionally, in section 17(1)(e) the provider of the social media service can remove a comment from an online page if the commenter consents to the removal (and if the comment was made in Australia). However, it needs to be clear that a social media provider is not restricted from taking action to remove material to protect itself from liability or for breach of terms of use.

Clause 22: Nominated entity of the provider of a social media service

Section 22 of the Bill proposes that if a social media service provider is a foreign body corporate, and the service has at least 250,000 Australian account-holders (or has been specified in the legislative rules), the provider must have a nominated entity in Australia. The nominated entity is required to have "access" to relevant contact information and country location data (s 22(1)(f)).

A nominated entity requirement as it currently stands would be unduly disruptive to existing data, privacy, and security arrangements. As a global company, Twitter operates in jurisdictions around the world and has



nominated personnel within select entities that have authorisation to access user data. We impose these strong restrictions in line with best practice for human resources, cyber-security considerations, and applicable law.

Additionally, this calls into question cross-border data transfer arrangements between countries under laws such as the European Union (EU) General Data Protection Regulation (GDPR) and EU-US Privacy Shield that determine if or how user data may be shared across borders and corporate entities. Notably, the requirement to have a nominated entity as defined under the Bill may also conflict with Australia's free trade obligations with the United States (US) relating to the provision of services. Article 10.5 of the Australian-United States Free Trade Agreement (AUSFTA) provides that Australia cannot require a service supplier in the US to establish or maintain a representative office or any form of enterprise, or to be resident, in Australia as a condition for the cross-border supply of a service.³⁰

It is also important to note that the current form of the Bill is drafted so that recourse for complainants are sought from, and obligations to comply, are placed on providers regardless of whether a nominated entity exists (e.g. the complaints mechanism (s 17) and avenue for complainants to seek end-user information disclosure orders against the provider (s 19).

The critical intent behind this section of the Bill is whether the complainant will have their complaint answered within the necessary timeframe; the complainant is not concerned with whether the answer is provided by a local entity. Therefore, the imposition of this requirement, as well as substantial penalties for failure to comply, including the continuing contraventions provision in section 23 of the Bill, is therefore an hostile result.

Addressing the High Court's decision in *Voller*

Twitter agrees that liability for publication of defamatory matter requires modernisation for the digital age, particularly in light of the decision in *Fairfax Media Publications Pty Ltd v Voller; Nationwide News Pty Limited v Voller; Australian News Channel Pty Ltd v Voller* [2021] HCA 27 ('*Voller*'). We also understand that the impact of the *Voller* decision has resulted in the need to reconsider and reallocate responsibility for defamatory material made online.

Allocating liability for publication to social media service providers ('providers') as contemplated in section 15(1)-(2) of the Bill, however, does not address the issues that *Voller* presents.

Potential liability for defamatory third-party material online should accrue where the page owner or provider is made aware of the publication, provided with sufficient information about the complaint, and fails to remove the publication within a reasonable time, which resonates with global concepts of safe harbour and innocent defamation defence. This suitably aligns with the mandatory concerns notice process required to commence proceedings under State and Territory legislation. Additionally, unlike the SMAT Bill, the Stage 2 Review of the MDPs considers potential orders to have online content removed, potentially addressing this gap.

Such a solution needs to take into consideration the balance of the protection of reputation to one's right to freedom of expression, and any amendment of the law in relation to liability for publication as affirmed in *Voller* needs to be workable and future-proofed for the digital age.

Conclusion

As drafted, the Bill functionally puts private companies in the invidious position of having to assess and determine the likelihood of a complainant's right to obtain relief through defamation proceedings, displacing the essential role and function of the courts. The net consequence would be to make applicable companies act as sole arbiters of truth with regards to online speech.

³⁰

<https://www.dfat.gov.au/trade/agreements/in-force/ausfta/australia-united-states-fta#:~:text=The%20Australia%2DUnited%20States%20Free,tariff%20lines%20went%20to%20zero.>



Twitter supports smart regulation, and our focus is on working with governments to ensure that regulation of the digital industry is practical, effective, and feasible to implement while remaining inclusive and keeping core democratic values intact while promoting tech innovation, including our core commitment to an Open Internet worldwide. In this vein, we support strong mechanisms to protect against defamation and assist in the swift removal of illegal content, but we emphasise the need to balance and protect principles of free expression to prevent a chilling effect on robust and open public discourse and avoid any unintended harmful consequences.

Going forward, we look forward to continuing to work collaboratively and in good faith with both the Federal, State and Territory Governments on this important area of the law, as well as partners in civil society, not-for-profits, academia, and industry to address online safety and work to create lasting global solutions to build a safer and Open Internet.