Meta's submission on the *Social Media (Anti-Trolling) Bill 2022*

Submission to the Senate Legal and Constitutional Affairs Legislation Committee

Meta

Executive summary

Meta welcomes the opportunity to provide input to the Senate Legal and Constitutional Affairs Legislation Committee's consideration of the *Social Media (Anti-Trolling) Bill 2022 [Provisions]*, the proposed "anti-trolling" legislation.

Although the commentary around this legislation has portrayed it as targeting online harassment, in reality, it relates to Australia's defamation laws. Meta has long supported reform of Australia's defamation laws to update them for the digital age. We supported the first stage of defamation reform agreed by Attorneys-General, and have been constructively engaging with the working process considering the second stage of defamation reform, led by New South Wales.

In our submission to the consultation process for the second stage of defamation reform, we said that the role of internet intermediaries in relation to defamatory content deserves further examination by Australian policymakers.[1] The New South Wales Government's discussion paper contained thoughtful and welcome discussion about the need to clarify and strengthen intermediary liability around defamatory content. We strongly support the position put forward at that time by Attorneys-General 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. The need for reform has been highlighted in recent cases which created significant uncertainty for internet intermediaries by essentially requiring them to block access to content, on the sole basis of a user allegation that the content is defamatory, rather than via a rigorous independent process.[2]

As well as clarifying the defences that internet intermediaries can rely on, we have also recognised that defamation law should enable efficient resolution for individuals who are the subject of potentially defamatory material.

We have no concerns with the apparent core policy objective of the legislation: we recognise that internet intermediaries can and should play a role in connecting complainants with the authors of posts that are potentially defamatory. A legislative requirement for companies to provide this information - subject to appropriate checks and balances - can be an effective way to enable resolution between the two parties, like the Norwich Pharmacal orders framework in the United Kingdom.

[1] Facebook Australia, *Facebook submission to the review of model defamation provisions (stage 2)*
[2] Defteros v Google LLC [2020] VSC219.

However, the legislation goes far beyond what is required to achieve this policy objective, in four areas which should concern parliamentarians, civil society and the Australian public.

1. **The legislation will embolden those people to make defamatory statements because they are now less likely to be held responsible.**

   The public rhetoric around the legislation has suggested it is intended to remove harassing or defamatory material online.

   However, the explanatory materials to the legislation reveal that it is the intended design of the legislation *not* to incentivise social media companies to remove more material than they currently do. It is intended to have a neutral impact (ie. not result in any greater takedowns than is currently the case).

   We believe the consequences of the law could go further and in fact embolden people to make defamatory statements online.

   Currently, these people would be primarily responsible and legally liable for the things they say online. This provides a strong incentive for people to be careful about what they say.

   However, the legislation makes the social media platform equally as responsible for the content as the author, and it removes social media services' defence of innocent dissemination. Given social media companies are almost always going to be more attractive defendants than the authors of the post, those people who may wish to make defamatory claims online will essentially be able to do so with impunity. The legislation will likely embolden them to post content which is directly harmful or which incites harmful comments, knowing that the social media company will be held responsible rather than the author.

   In essence, the legislation assigns full liability (in a practical sense) for all online content to the social media platform. This is even though the platform has no editorial control over the content of posts or context to the claims and may not even be aware of the post's existence.

2. **It creates inequity in relation to defamation complaints, based on the type of digital platform.** It represents a Commonwealth takeover of the provisions of defamation law that apply to social media platforms, while maintaining the

existing defamation regime at the state and territory level for all other platforms, publishers and authors. This means a defamatory review posted on an ecommerce website or a defamatory comment on a news publisher's own website will be subject to a different liability regime than if the same content was posted on a social media service. Case law is continuing to develop through the courts and splitting the law in this way risks greater complexity and divergence in the requirements faced by different companies.

3. **The legislation undermines Australia's efforts to demonstrate global leadership in encouraging cross-border data flows and an open internet.** The legislation establishes new local presence requirements, and compels changes to social media companies' corporate structure and local operating requirements. As a principle, we strongly disagree with regulations that prescribe the specific corporate structure that a social media platform should take in Australia. It is also not necessary for securing access to the relevant user information. Given Meta's existing corporate structure facilitates the Government's desired policy outcomes, we see no valid reason why this legislation should necessitate this change.

   Additionally, Australia has previously played a vital role in arguing against forms of localisation around the region. For Australia to contemplate a requirement for technology companies to establish a local presence sets a concerning precedent that could undermine the principles of an open internet and embolden other countries with a different vision of the internet's future.

4. **The legislation will lead to increased collection of data about Australians, including by companies based in China.** It incentivises much greater collection and verification of Australians' data by social media platforms, most of whom are headquartered overseas. Social media platforms are not able to access the legal defence unless they can provide relevant and current contact details for any Australian end-user accused of defamation, and demonstrate that contact will be successful on the basis of those contact details. These requirements go beyond what is needed to pursue litigation against a person, or to serve legal documents.

Any of these new requirements individually would be a major change to Australians' experience of online social media services. Collectively, the consequences could be very significant, from the perspective of defamation, privacy, and the global contest of differing visions for the internet.

As with our submission to the Attorney-General's Department, we provide a number of constructive suggestions in this submission to address the practical issues arising from the legislation. We believe significant amendments are required before the legislation would be workable or able to effectively achieve the policy objective of reducing harmful content online. If it is rushed through Parliament in order to meet an election timetable, there could be serious consequences for Australians' use of social media products as digital platforms work to notify millions of Australians of the need to share and regularly verify their personal contact details online. Instead, we urge policymakers to take the time to ensure defamation reform is effective, proportionate and durable.

Recommendations

State and territory Attorneys-General have conducted a lengthy and detailed review of the model defamation law over the past eighteen months. They invited and considered feedback from interested parties and Meta provided constructive feedback on the reform. The Stage 1 reforms have been implemented in most states and territories.

Our primary recommendation is that Australian policymakers do not rush this legislation. We urge the Parliament to take the time to carefully consider the possible unintended consequences - including the interaction with other, recently-reformed laws - and work with all relevant stakeholders to set a framework for defamatory content online that can be effective, proportionate and durable. It is not possible to see how a workable piece of legislation could be passed before the 2022 federal election.

We have provided some specific recommendations about how the concerns with the legislation could be addressed:

1. Social media services should be able to continue accessing the defence of innocent dissemination, similar to all other secondary publishers and internet intermediaries. Section 15(3)(f) should be removed.

2. Requirements for social media services to operate a nominated entity in Australia should be removed, as the draft legislation operates outside Australia and providers are required by Australian law to comply with end-user information disclosure orders. Or, at the very least, this requirement should be adjusted so that it is only required of those service providers who do not have established processes for Australian defamation complainants to follow.

3. At a minimum, a complaints process defence should be workable and set a reasonable standard that is possible for social media services to meet.

   For instance, a social media service provider should be able to rely on a complaints scheme defence if:
   (1) the applicant already knows the poster's relevant contact details, or is reasonably able to ascertain the details on their own;
   (2) the social media service provider promptly removes the material on notice or the poster removes the material;
   (3) the applicant has requested but the court has not made, or refuses to make, an end-user information disclosure order;

(4) the social media service provider complies with an end-user information disclosure order (regardless of the scope of information ordered to be disclosed under that order);
(5) the social media service provider reasonably believes taking action under its complaints scheme may present a risk to the poster's safety; or
(6) the social media service provider reasonably believes that the complaint or the request does not genuinely relate to the potential commencement by the complainant of defamation proceedings against the poster in relation to the material.

4. In order to ensure any data collection that occurs as a result of the legislation is proportionate and necessary, the definition of "relevant contact details" should be amended.

   The definition should be limited to the name and email address, or name and telephone number, held by the social media service provider. This is enough to enable substituted service of legal documents. Meta would support any changes to the substituted service rules that might be necessary.

   Further, a social media service provider should not be required to disclose contact details or country location data if it would breach the laws of another country.

5. Additional amendments should be made to ensure the scheme can operate effectively in practice:
   - **Alignment with concerns notice process:** For consistency and to avoid increasing litigation, the threshold for lodging a complaint should align with the concerns notice process, including satisfying the serious harm test. The onus of establishing that should continue to rest with the complainant.
   - **Clarification of scope:** The change from "comment" to "material" should be reversed given the potential for such a change to significantly increase the scope of the Bill. The definition of "social media service" should be clarified to clearly exclude "relevant electronic services" as defined in the *Online Safety Act 2021* (Cth). In addition, the legislation should apply to defamation proceedings which directly "concern" material (and not those that simply "relate to" material).
   - **Response times:** The timeframes should be extended to a reasonable period, or there should be some flexibility built into them, to allow for a provider to request additional information from a complainant, to liaise with

a poster who is slow to respond or to consider difficult cases (such as where there could be a risk of safety to the poster).

6. **Clarity on process for making complaints:** Social media service providers should be able to specify a single point or channel in which complaints are made. This could include, for example, a single email address or digital complaints form, to allow providers to effectively manage the unreasonably short turnaround times. The legislation should state a complaint is "made" when received by the social media service provider at that channel. While the Government has introduced a right for it to prescribe legislative rules in relation to how a provider is required to communicate with a complainant, there is no limitation on how a complainant is required to communicate with a provider.

Table of contents

Meta's existing work

Relationship between defamation and online safety

Much of the commentary and explanatory materials relating to the legislation have portrayed its intent and effect as being to combat online harassment, particularly of women and young people. This focus is also reflected in the inclusion of "anti-trolling" in the legislation's name. In reality, however, the legislation relates to Australia's defamation laws.

Defamation law should not be equated with online safety law. Australian defamation law is designed to protect the reputation of an individual (often a public figure) by providing compensation for damage caused to their reputation. Defamation is not the same as trolling or harassment. Sometimes potentially defamatory claims are made in the public interest: in order to hold a public figure to account; to blow the whistle on corruption or unethical behaviour; or to criticise an individual in a position of power. Defamation law recognises this and allows certain defences to apply, such as public interest or qualified privilege. On the other hand, online safety laws are designed to minimise the harms associated with online harassment and trolling. Defamation and online safety laws serve different purposes and care should be taken to keep them separate.

We are committed to working constructively with the Australian Government on setting effective regulatory frameworks to hold companies such as Meta to account and set rules for the internet.

While we support reform of Australia's defamation laws for the digital age, conflating defamation and safety law confuses the issues and ultimately inhibits the ability of policymakers to design effective law reform.

The Australian Government is currently undertaking significant online safety and privacy reform processes, and we respectfully suggest that any regulatory frameworks targeted at online harassment and trolling, particularly of women and young people, should form part of the online safety reforms that are currently underway, and not form part of defamation reform. The recently enacted *Online Safety Act 2021* (the Act), which just took effect on 23 January 2022, together with the Basic Online Safety Expectations and the code development work currently underway to implement the Act, are more suited to address the harms associated with online harassment than defamation law. Unlike defamation law, which is focussed on compensation for damage to reputation, the Act is focussed on harm minimisation through the removal of serious and targeted online

abuse.[3] The eSafety Commissioner has the power to require the removal of such content within 24-hours and to require the disclosure of contact details for a user to support enforcement action against that user.

Work to combat online harassment

To understand the context of online safety, we have outlined Meta's work to combat online harassment and protect people online.

To combat harmful online harassment and trolling, we: have policies that are designed to keep people safe on our services; invest in operational teams and technology to enforce these policies; and develop tools to assist people to take further action to prevent online harassment. We also develop resources and are grateful for partnerships with many Australian child safety, mental health, women's services and other organisations to promote awareness of our policies, tools and resources.

Policies

Our policies, known as our Community Standards,[4] outline what is and is not allowed on Meta's services. These policies are developed based on a range of values to help combat abuse. Safety is a core value of our Community Standards, alongside privacy, authenticity, voice, and dignity.[5]

Our Community Standards prohibit various categories of harmful content, including – most relevantly to online harassment and anonymous trolling – prohibiting fake accounts as well as bullying and harassment.

Our policies are based on feedback from our community, and the advice of experts in fields such as technology, public safety, child safety and human rights. To ensure that everyone's voice is valued, we take great care to craft policies that are inclusive of different views and beliefs, in particular those of people and communities that might otherwise be overlooked or marginalised.

We regularly update our policies to reflect society's expectations and feedback from experts and stakeholders. We made a major update to our policy around bullying and

[3] *See* https://www.esafety.gov.au/newsroom/media-releases/safety-net-protect-australian-adults-serious-online-abuse-2022
[4] See Meta, *Community Standards*, https://www.facebook.com/communitystandards
[5] Monika Bickert, *Updating the values that inform our community standards*, https://about.fb.com/news/2019/09/updating-the-values-that-inform-our-community-standards/

harassment (particularly of public figures) in November 2021.[6] In recent years, we also updated our policies to adjust for the gendered and culturally specific nature that some forms of online harassment and abuse can occur, especially for women. In July 2019, our policy team expanded our bullying and harassment policy to enforce more strictly on cursing that uses female-gendered terms.[7]

Enforcement

In order to enforce our policies, we invest very significantly in both technology and people to help detect violating content or suspicious behaviour. In relation to "anonymous trolling", we have a high rate of proactive detection of fake accounts and are rapidly increasing our ability to do this for bullying and harassment.

We have built up teams of experts who work in this space. We now have over 40,000 people dedicated to keeping people safe on our apps. We've invested more than US$13 billion (~AU$18 billion) on safety and security since 2016, and we spent more than US$5 billion (~AU$6.9 billion) in 2021.

We encourage users to report content that they are concerned about. Once reported, we assess these reports and action the content consistent with our policies. However, increasingly, we have been investing in proactive detection technology to identify and action harmful content such as fake accounts and bullying and harassment before anyone sees it and needs to report it to us.

We have scaled our enforcement to review millions of pieces of content across the world every day, and use our technology to help detect and prioritise content that needs review. To provide transparency to the community that can be used to hold us to account, we provide data about our enforcement work with respect to our global policies in our Community Standards Enforcement Report.[8] The report is released quarterly and includes metrics such as how much content we are actioning, and what percentage was detected proactively. We now report on 14 policy areas on Facebook and 12 on Instagram.

---

[6] A Davis, Our approach to addressing bullying and harassment, *Meta Newsroom,* 9 November 2021, https://about.fb.com/news/2021/11/how-meta-addresses-bullying-harassment/

[7] Meta, Making Facebook a safer, more welcoming place for women, *Meta Newsroom,* 29 October 2019, https://about.fb.com/news/2019/10/inside-feed-womens-safety/

[8] Meta, *Community Standards Enforcement Report* https://transparency.fb.com/data/community-standards-enforcement/

From July to September 2021, our latest Community Standards Enforcement Report confirms that we removed 1.8 billion fake accounts from Facebook, 99.8% of which we actioned proactively before it was reported to us.[9]

In relation to bullying and harassment material, in the third quarter of 2021:[10]

- We actioned 9.2 million pieces of content on Facebook for violating our policies on bullying and harassment, and of that, 59.4 per cent of bullying and harassment content was removed proactively via artificial intelligence. This is an increase from 54.1 per cent in the previous quarter, and 25.9 per cent one year prior.
- We actioned 7.8 million pieces of bullying and harassment content on Instagram, and of that, 83.2 per cent of it was removed proactively. This is an increase from 71.5 percent in the previous quarter and 54.5 per cent one year prior.

Tools

We build technology to help prevent abuse and harmful experiences such a misuse of our services by fake accounts or from bullying or harassment in the first place, and we also design tools to give people more control and help them stay safe. We believe people should have tools to customise their experience on our services - even if content does not violate our policies, people may still find it objectionable or may choose not to see it.

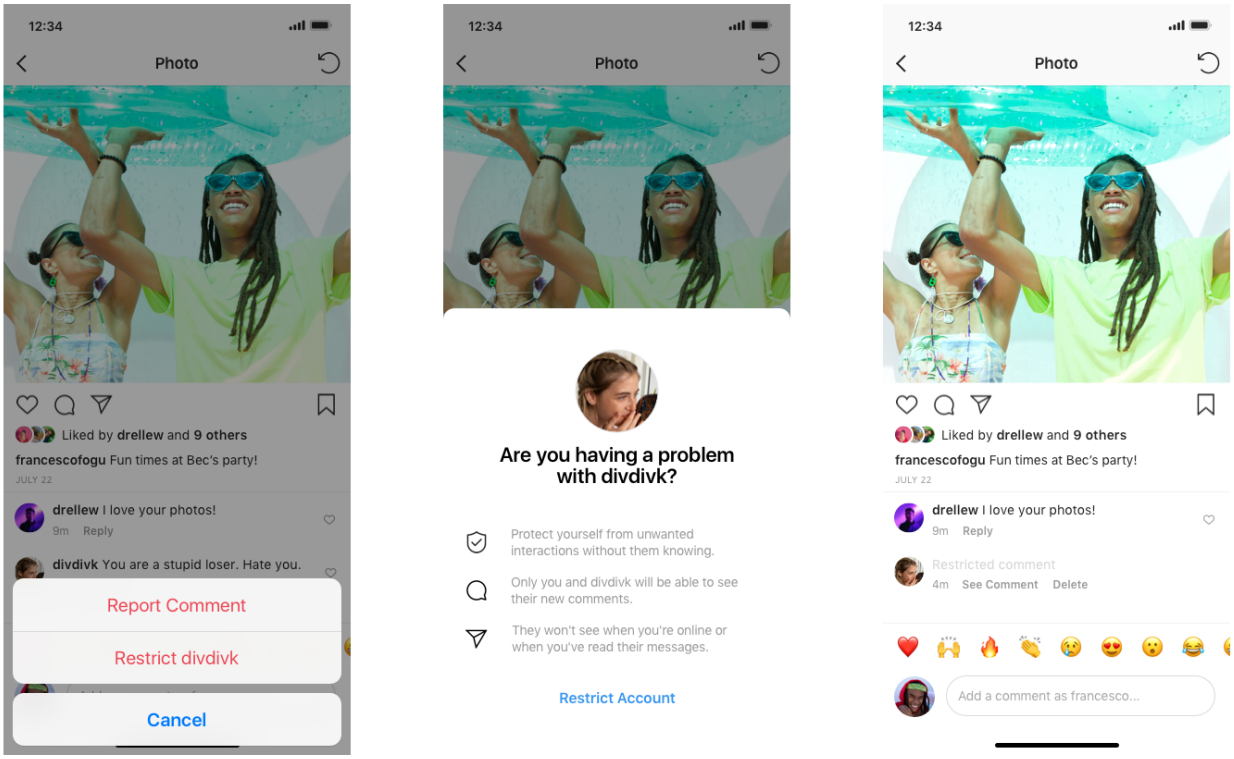
In addition to the long-standing tools of Block, Report, Hide, Unfollow,[11] we continue to introduce new features to help users manage their experience. These tools are informed by our consultations with industry, experts and civil society organisations. Our tools aim to discourage harmful behaviour, particularly online harassment, and help users control their experience. Recent tools that we have released to help people combat online harassment and trolling include:

- **Restrict tool.** We've created a Restrict tool in Instagram[12], shown in Figure 1 below, where comments on your posts from a person you have restricted will only be visible to that person. Direct messages will automatically move to a separate Message Requests folder, and you will not receive notifications from a restricted account. You can still view the messages but the restricted account will not be able

[9] *See* Meta, Community Standards Enforcement Report https://transparency.fb.com/data/community-standards-enforcement/fake-accounts/facebook/
[10] Meta, *Community Standards Enforcement Report Q3 2021 - Bullying and harassment,* https://transparency.fb.com/data/community-standards-enforcement/bullying-and-harassment/facebook/
[11] An overview of these and other tools is available in the Facebook Safety Center: https://www.facebook.com/safety/tools
[12] Instagram, 'Introducing the "Restrict" Feature to Protect Against Bullying', *Instagram Blog*, 2 October 2019, https://about.instagram.com/blog/announcements/stand-up-against-bullying-with-restrict.

to see when you've read their direct messages or when you are active on Instagram. This means you can protect yourself from harmful messages without triggering further abuse from a person, if they become aware you have blocked them.

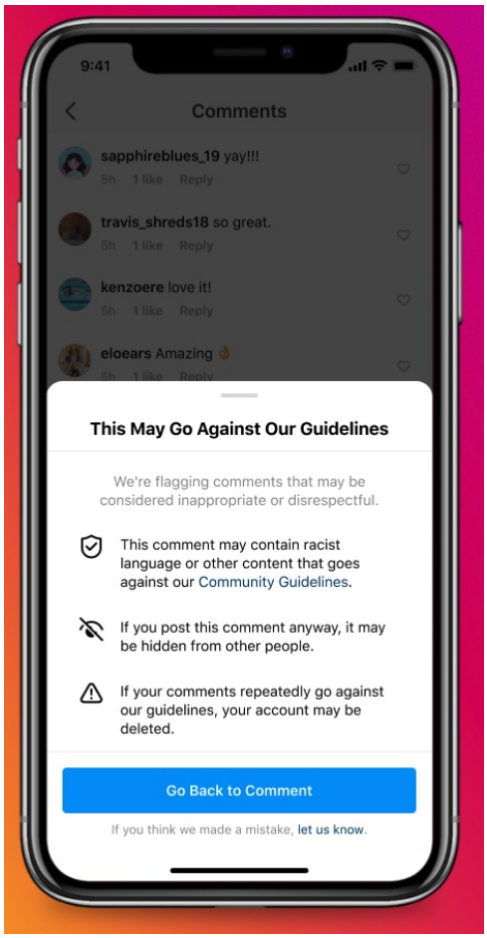
Figure 1: Instagram 'Restrict' tool



- **Bullying and harassment warning**. One recent tool we've deployed on both Facebook and Instagram is sending a warning to educate and discourage people from posting or commenting in ways that could be bullying and harassment, shown in Figure 2 below. We found that after viewing these warnings on Instagram, about 50 per cent of the time the comment was edited or deleted by the user.[13]

[13] A Davis, Our approach to addressing bullying and harassment, *Meta Newsroom,* 9 November 2021, https://about.fb.com/news/2021/11/how-meta-addresses-bullying-harassment/

Figure 2: Warnings to discourage bullying or harassment

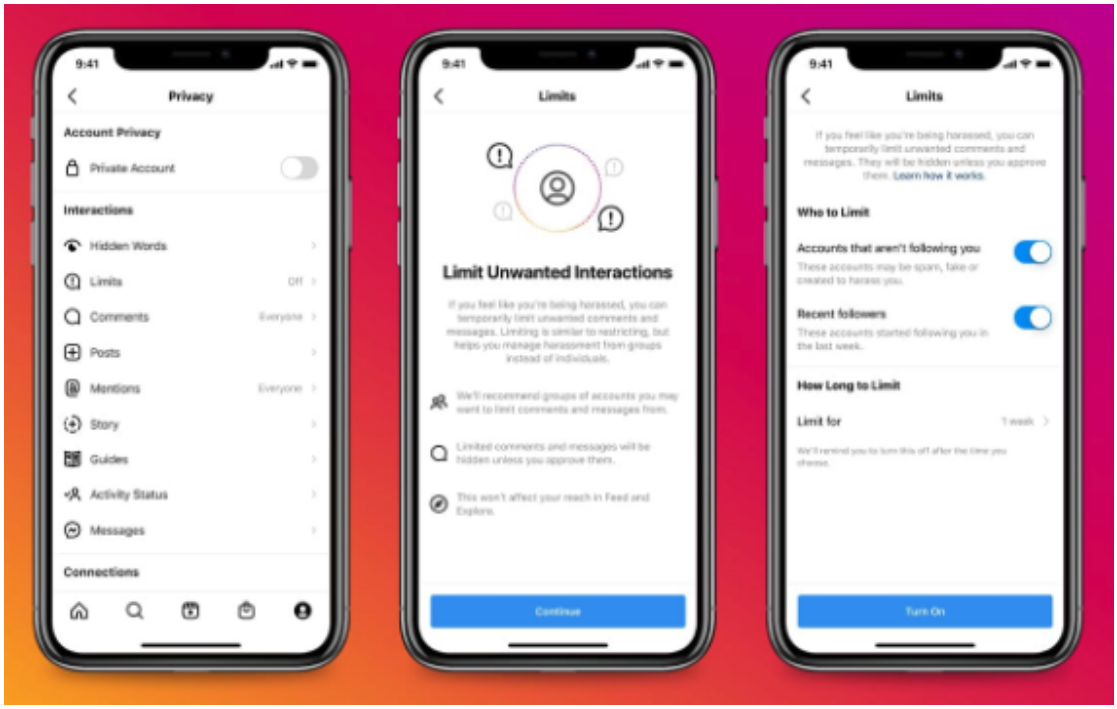


In consultation with experts and public figures themselves, we have introduced a number of specific tools that help users reduce unwanted interactions online, including:

- **Limits.** The Limits tool on Instagram, shown in Figure 3 below, allows users to automatically hide comments and Direct Messaging requests from people who don't follow them, or who only recently followed them, to help manage an unexpected rush of unwanted contact.[14] We developed and launched this tool in partnership with the Australian Football League (AFL), to help protect their players from racist abuse. This tool is particularly useful for public figures to protect them from trolling.

[14] Instagram 'Introducing New Ways to Protect our Community from Abuse', *Instagram Blog,* 10 August 2021, https://about.instagram.com/blog/announcements/introducing-new-ways-to-protect-our-community-from-abuse

Figure 3: 'Limits' tool on Instagram



- **Control who can comment on Facebook.** In March 2021 we introduced new tools to give users more control over who can comment on their posts on Facebook News Feed. Users can control their commenting audience for a public post by choosing from a menu of options. By adjusting the commenting audience, users can further control how they want to invite conversation onto their public posts, and limit potentially unwanted interactions.[15]

- **Comment Controls on Instagram.** The Comment Controls feature on Instagram allows users to automatically hide comments based on a list of words, phrases, numbers or emojis that they can manually add to based on their experiences or preferences.[16] If people comment using those words or emojis, the user will not be notified and they will not be published on the post for anyone to see. We know from research that, while people don't want to be exposed to negative comments, they want more transparency into the types of comments that are hidden. You can tap "View Hidden Comments" to see the comments. Comments that violate our

[15] R Sethuraman, More control and context in News Feed, *Meta Newsroom*, 21 March 2021, https://about.fb.com/news/2021/03/more-control-and-context-in-news-feed/
[16] Instagram, Kicking Off National Bullying Prevention Month With New Anti-Bullying Features, *Instagram Blog,* 6 October 2020, https://about.instagram.com/en_US/blog/announcements/national-bullying-prevention-month.

Community Guidelines will continue to be removed.

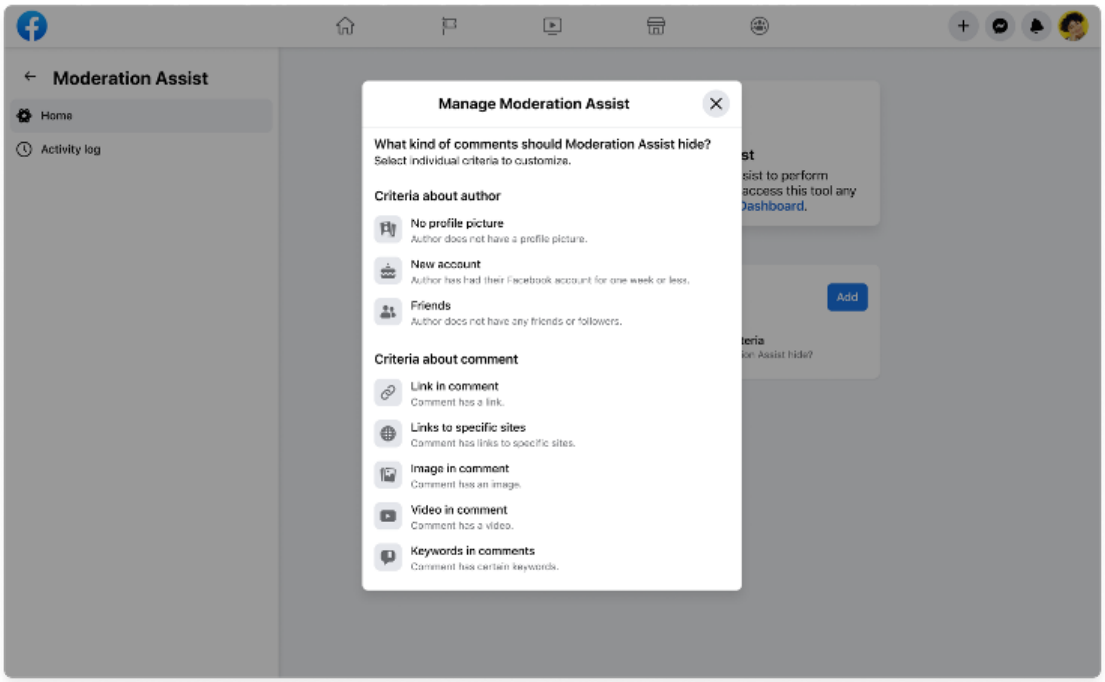
- **Hidden Words.** We have recently introduced a tool which will automatically filter direct message (DM) requests containing offensive words, phrases and emojis.[17] This tool focuses on DM requests, because this is where people usually receive abusive messages - unlike the regular DM inbox - where you receive messages from friends. We have worked with leading anti-discrimination and anti bullying organisations to develop a predefined list of offensive terms that will be filtered from DM requests. Users also have the option to create their own custom lists of words, phrases or emojis that they don't want to see in their DM requests, because we know that different words can be hurtful to different people.

- **New blocking features.** To protect users from unwanted contact, last year we launched new blocking features so that whenever you decide to block someone on Instagram, you'll also have the option to block new accounts that person creates.[18] This is designed to prevent users from being contacted by someone they've blocked, even when they create a new account. This is in addition to our harassment policies, which already prohibit people from repeatedly contacting someone who doesn't want to hear from them.

- **Moderation assist.** In February 2022, we introduced Moderation Assist for New Pages Experience on Facebook.[19] The tool allows Page admins to easily set moderation criteria for all posts, reducing how much time is spent on comment moderation. Admins can select from a list of criteria to automatically hide certain types of comments, such as comments with links, from profiles with no profile photo, from new followers, and more. Admins can control the criteria with the ability to undo any actions or turn it off at any time.

[17] Instagram, Introducing new tools to protect our community from abuse, *Instagram Blog*, 21 April 2021, https://about.instagram.com/blog/announcements/introducing-new-tools-to-protect-our-community-from-abuse
[18] Ibid.
[19] Facebook, What is moderation assist for Facebook pages?, *Facebook Help Centre,* https://www.facebook.com/help/1011133123133742

Figure 4: Moderation assist



Resources

Fourth, we provide informative resources and learning modules for our users to raise awareness of online safety, and the tools available to help them manage their experience. This includes the Instagram Safety and Wellbeing Hub[20] and the Facebook Safety Center.[21] These also include the:

- Bullying Prevention Hub developed in partnership with the Yale Centre for Emotional Intelligence;
- Youth Portal which provides a central place for teens to access education on our tools and products, first person accounts from teens about how they're using technologies, tips on security and reporting, and advice on how to use social media safely; and[22]
- Get Digital Hub, a digital citizenship and wellbeing program which provides schools and families with lesson plans and activities to help build the core competencies and skills young people need to navigate the digital world in safe ways.[23]

[20] Instagram, *Instagram Community*, https://about.instagram.com/community
[21] Meta, *Digital Literacy Library*, https://www.facebook.com/safety/educators
[22] Meta, *Youth Portal*, https://www.facebook.com/safety/youth?locale=en_GB
[23] Meta, *Get Digital Hub*, https://www.facebook.com/fbgetdigital

Partnerships

Finally, we have over 400 safety partners across the world, including a number of partnerships in Australia, to ensure our global safety efforts are complemented by on-the-ground expertise.

Globally, we have a Safety Advisory Board, which comprises leading safety organisations and experts from around the world. Board members provide expertise and perspective that inform Meta's approach to safety. The Australian youth anti-bullying organisation PROJECT ROCKIT is one of 11 organisations globally that serves on this Board.

In 2020, we were also one of the first technology companies to appoint a Global Head of Women's Safety, and in 2021 we announced our Global Women's Safety Expert Advisors,[24] a group of 12 nonprofit leaders, activists and academic experts to help us develop new policies, products and programs that better support the women who use our apps. This expert group includes Dr Asher Flynn, an Associate Professor of Criminology at Monash University and the Vice President of the Australian and New Zealand Society of Criminology.

In Australia, we invest significantly in local organisations to promote important safety and wellbeing messages. For example, we have invested in a Digital Ambassadors program delivered by PROJECT ROCKIT.[25] Digital Ambassadors is a youth-led, peer-based anti-bullying initiative. A Digital Ambassador aims to utilise strategies to safely connect and tackle online hate. This is a nine-year partnership that has directly empowered more than 25,000 young Australians to tackle cyberbullying.[26]

We have also developed an Australian Online Safety Advisory Group to consult and provide a local perspective on policy development. This group comprises experts such as CyberSafety Solutions, PROJECT ROCKIT, WESNET, and the Alannah and Madeline Foundation, as well as many others.

We actively promote awareness of our policies, enforcement, tools and resources by partnering with influential Australian stakeholders; within the last six months, for example, we have:

- Hosted an online discussion in September 2021 with Minister for Communications Paul Fletcher, Dr Asher Flynn and Cindy Southworth, Global Head of Women's

[24] C Southworth, Partnering with experts to promote women's safety, *Meta Newsroom*, 30 June 2021, https://about.fb.com/news/2021/06/partnering-with-experts-to-promote-womens-safety/
[25] Project Rockit, *Launching: Digital Ambassadors,* https://www.projectrockit.com.au/digitalambassadors/
[26] R Thomas, 'Young People at the Centre', *Facebook Australia blog*, 8 February 2021 https://australia.fb.com/post/young-people-at-the-centre/

Safety Safety Policy at Meta, together with Mamamia, to raise the profile of work being done by Government, industry and community to support women's safe experience online.[27] The event reached more than 32,000 people.

- Participated in a panel event, also in September 2021, for the Parliamentary Friends of Making Social Media Safer alongside the eSafety Commissioner, in order to help raise awareness among Australian Parliamentarians about tools that are available to help keep them safe online.
- Worked closely with sporting organisations such as the AFL. Most recently, in December 2021, Meta worked with the AFL to deliver a specialised education workshop for AFL men's and women's players to understand the tools and resources available to them and provide an additional layer of support through peak season moments. This workshop included participation by the eSafety Commissioner's Office.

[27] Meta, Women's Safety Panel, *Facebook*, 22 September 2021, https://www.facebook.com/watch/?extid=NS-UNK-UNK-UNK-IOS_GK0T-GK1C&v=266923491951411

Meta's current approach to defamation

As mentioned above, our Community Standards are a global set of policies that outline what is and is not allowed on Meta's platforms. Given the global and diverse nature of the community we serve, our Community Standards do not reflect any specific legal system, nor are they intended to cover all local laws. However, because they are designed to prevent harm, they do overlap with local law in a number of instances.

In addition to reviewing content against our Community Standards, we also consider whether content reported to us is lawful under local law (such as Australian defamation law). A person may report content which they believe is defamatory under local law using our dedicated defamation reporting form.[28] We have a team of trained lawyers and subject matter experts who handle defamation reports submitted by Australian users, and review the reported content against Australian defamation law. Where the content is clearly unlawful, we restrict access to the content in Australia, unless such action would be inconsistent with international human rights standards or have an unduly adverse effect on the availability of speech via our products.

Evaluating defamation reports presents unique considerations for intermediaries like Meta. While we consider Australian defamation law when we are alerted to content that is allegedly defamatory in Australia, Meta: (a) does not exercise editorial control over the content that is distributed on its services (our users have that control); and (b) often does not have sufficient information to assess whether the content is clearly unlawful, including whether relevant defences (such as truth) are likely to apply.

Following recent cases in Australia, internet intermediaries have been essentially required to block access to content, on the sole basis of a user allegation that the content is defamatory, rather than via a rigorous independent process.[29] This presents a very real risk of erroneously over-blocking content in an attempt to bridge this knowledge gap and avoid liability, which could have a chilling effect on international human rights such as freedom of expression.

If a person who has reported the content wishes to serve defamation proceedings on the poster of the content, but cannot identify the poster of the content based on the information available to them, the reporter can apply for a pre-action discovery order for disclosure of the poster's contact details. Meta complies with such an order made in Australia, provided that such disclosure is permitted by applicable law and our terms of service, and the data is available and reasonably accessible. The US location of Meta

[28] *See* https://www.facebook.com/help/contact/430253071144967
[29] Defteros v Google LLC [2020] VSC219.

Platforms, Inc. is not in itself a barrier to the provision of data pursuant to a valid Australian court order.

Meta made detailed submissions on the Council of Attorneys-General's Stage 1 and Stage 2 discussion papers for the review of the Model Defamation Provisions. As noted in those submissions, the role of internet intermediaries in relation to defamatory content deserves further examination by Australian policymakers and we expressed our view that Australian laws are in many ways not fit-for purpose in the digital world.

We also expressed our support for clarification of the application of the innocent dissemination defence to intermediaries, as well as the introduction of a process designed to connect complainants with the originators of content. These are complex issues, which have benefitted from the thoughtful consideration given to them by the Stage 2 reform process, led by Attorneys-General.

Finally, while Meta does not have editorial control over the content distributed on our services, we provide users in Australia with tools to limit the risk of defamatory material on their Page or Group. For example, as mentioned above, in March 2021, we released a new product called Control Who Can Comment, which allows the admin of a Facebook Page to control who can comment on particular posts. We also launched a new tool in February 2022, Moderation Assist, which allows admins to set moderation criteria that will automatically hide certain types of comments.[30] The ability to limit comments directly empowers Facebook Page admins to manage third party comments and potential liability arising from decisions like *Voller*.

Although the decision of the High Court in *Voller* rightly raised questions about whether defamation laws were workable for the internet age, the decision was in line with common law precedent for publication in Australia.[31] No defences have yet been raised and therefore there are open questions about whether defences (such as the innocent dissemination defence) may be successful in this case. Similarly, recent and forthcoming technology and product developments - such as the Control Who Can Comment product - may provide avenues for managing the potential liability of Page owners in a manner more proportionate than wholesale excluding them from any possible liability.

[30] Facebook, What is moderation assist for Facebook pages?, *Facebook Help Centre,* https://www.facebook.com/help/1011133123133742
[31] *Webb v Bloch* (1928) 41 CLR 331; *Trkulja v Google LLC* (2018) HCA 25

Concerns with the legislation

The legislation will embolden people to post defamatory material on social media

The public rhetoric around the legislation has suggested it is intended to remove harassing or defamatory material online.

However, the explanatory materials to the legislation reveal that it is the intended design of the legislation *not* to incentivise social media companies to remove more material than they currently do. It is intended to have a neutral impact (ie. not result in any greater takedowns than is currently the case).

We believe it could go further and in fact embolden people to make defamatory statements online, especially as complainants seek to test the application of the regime.

Currently, these people would be primarily responsible and legally liable for the things they say online. This provides a strong incentive for people to be careful about what they say.

However, the legislation makes the social media platform equally as responsible for the content as the author, and it removes social media services' defence of innocent dissemination.

Given social media companies are almost always going to be more attractive defendants than the authors of the post, those people who may wish to make defamatory claims online will essentially be able to do so with impunity. As currently drafted, and contrary to the objective of focusing the dispute between "originator and victim", the legislation incentivises complainants to file proceedings against social media service providers rather than the authors or originators.

The consequence is that the legislation will embolden them to post content which is directly harmful or which incites harmful comments, knowing that the social media company will be held responsible rather than the author. It is likely to encourage increased harmful content online, increased defamation complaints, and fail to deter those perpetrating online harassment as it: (1) removes potential liability of page owners, and thus their incentive to share thoughtfully and moderate content; (2) makes social media service providers liable as publishers from the time that content is posted (including

before they are aware of it); and (3) does not provide a defence where content is removed by either the poster or by the provider.

In essence, the legislation assigns full liability (in a practical sense) for all online content to the social media platform. This is even though the platform has no editorial control over the content of posts or context to the claims and may not even be aware of the post's existence.

Secondly, the regime proposed by the legislation is open to abuse by complainants, who may see it as an opportunity to bring deep-pocketed defendants into proceedings. For example, a complainant may know, or may be able to easily ascertain, the contact details of a user, but may still bring proceedings against the social media provider. This allocation of liability does not seem consistent with the objective of focusing legal proceedings between the victim and the originator of the comment.

The law undermines Australia's global and regional leadership on data flows and could lead to internet fragmentation

The legislation contains requirements for social media service providers to establish nominated Australian entities that can access user data for users who have posted material while in Australia. The rationale for this requirement is that it addresses potential pragmatic and jurisdictional matters that could present a barrier to the operation of the complaints and end user information disclosure order mechanisms.

This rationale is not supported by the evidence of Meta's experience and processes for responding to defamation complaints in Australia, and is not warranted with respect to Meta's family of apps. Meta Platforms, Inc. (located in the US) can and does take action in response to Australian defamation reports and, just as contemplated in the legislation, Meta Platforms, Inc. may provide the available contact information of an end-user as permitted by applicable law, in accordance with our terms of service, and if the data is available and reasonably accessible.

As a principle, we strongly disagree with regulations that prescribe the specific corporate structure that a digital platform should take in Australia. Given Meta's existing corporate structure facilitates the Government's desired policy outcomes, there is no valid reason

why this legislation should necessitate the creation of a particular Australian entity with a particular defined legal relationship with Meta Platforms, Inc.

Additionally, for a country like Australia to contemplate a requirement for technology companies to establish a local presence sets a concerning precedent around the world. Local presence laws have been pursued in countries such as Vietnam, Brazil and Turkey in order to facilitate the surveillance or censorship of citizens' online activities, and violate individuals' human rights such as freedom of expression and privacy. If Australia pursues this approach, it may embolden other countries to follow this path, and lead to further internet fragmentation which will undermine the open internet.

This legislation would run counter to the objectives of encouraging cross-border data flows and an open internet that Australia has been pursuing in foreign affairs and trade policy as part of its work on e-commerce and digital trade through the World Trade Organization and multiple bilateral and multilateral trade agreements.

It is also inconsistent with some of Australia's existing trade agreements (such as the US-Australia Free Trade Agreement, which expressly prohibits Australia from requiring that US service suppliers, such as Meta, establish or maintain a representative office in Australia as a condition for the cross-border supply of its services to the Australian public)[32]. To rely on technical exceptions to justify the view that this legislation is compliant with the Free Trade Agreement would be seen as hypocrisy by other countries throughout the region.

We encourage Australian policymakers to consider the legislation against the broader geo-political context and state of the global internet.

The origins of the global internet were founded on liberal, democratic principles. An open internet has been pioneered by companies from the US - one of Australia's closest allies - and has enabled Australians to connect and small Aussie businesses to thrive. However, the values that underpin the original global internet are increasingly being challenged by a different vision of the internet pioneered by other countries – a heavily surveilled and closed internet with data localisation, and very little individual privacy.

[32] *See,* Article 10.5, United States-Australia Free Trade Agreement

This is why Meta has been calling for a "Bretton Woods" moment for the internet[33] – the creation of a multilateral, international framework for the internet that would agree the inviolable principles of how the global internet operates, These could include privacy of the individual, user rights, open data flows across borders, transparency and accountability by which the systems are operated, strict limits on the amount of sort of intrusive censorship, and agreement on whether governments or industries.

Australia should lead by example and implement domestic laws that are consistent with their stated policy agendas in multilateral international fora.

For this reason, we suggest that this requirement is removed or, at the very least, adjusted so that it is only required of those service providers who do not have established processes for Australian defamation complainants to follow.

Concerns over local presence requirements

We are are concerned that increased efforts around the globe to establish local presence requirements will ultimately and collectively harm economies and SMEs, and may lead to significant human rights concerns.

The internet enables companies to go global before or at the same time as they go national by leveraging e-commerce platforms, social media, and websites to find suppliers, customers, and partners anywhere in the world - thanks to economies of scale and supply chains that previously did not exist.

This has opened the door to a thriving global marketplace of opportunity for entrepreneurs to go beyond the limitations of a traditional brick-and-mortar operation because online platforms do not have to have a local presence in order to serve a local market with their goods and services.

What makes the digital revolution so powerful is the ability for businesses and consumers to provide and access products, services and information from anywhere in the world at low cost, and thus ensuring that businesses of all sizes and in all countries can access the technology and resources necessary to succeed.

---

[33] N Clegg, 'A Bretton Woods for the digital age can save the open internet', *Australian Financial Review*, 16 November 2021, https://www.afr.com/technology/a-bretton-woods-for-the-digital-age-can-save-the-open-internet-20211115-p5994h

Today, information and communications technology products and services act as both the catalyst and platform for small business growth and enhanced participation in international trade. The global reach of the internet gives way to easy communication and access to business partners, customers, information, and collaborators in a way that was never possible before.

Globally, local presence requirements:

- Limit consumer access to technology. Instituting additional barriers for setting up and running local offices and legal entities in each country can limit some countries from benefitting from products, services and information that they would otherwise have been able to enjoy.
- Restrict growth of SMEs and stifling innovation. Local presence requirements may limit small businesses' ability to engage potential partners and resources.
- Go against international trade norms. By instituting local presence requirements, countries are deviating from established international trade norms and practices that promote free trade by erecting unnecessary barriers to cross-border services trade. Furthermore, there is a risk that countries may reciprocate and impose similar requirements - impacting the growth of both local and international SMEs because requirements for local presence raise the costs for SME businesses that may be seeking to enter a new market. For example, a 2015 study by Leviathan Security Group found that localisation requirements raise the cost of businesses (potential costs of hosting data as well) by 30-60 percent.[34] This practice is reflected through international trade agreements, including the CPTPP, RCEP, and the AUSFTA.
- Poses non-tariff barriers to trade. Requiring local incorporation and presence unnecessarily discriminates against foreign businesses, and poses a non-tariff barrier to trade.
- Embolden other states to pursue similar requirements. The requirement of a physical establishment and local appointees may embolden other states to introduce local presence requirements paired with data localisation policies in order to provide easy access to data for law enforcement purposes. Personnel and data localisation measures such as those in India, Vietnam, Turkey and China, are often intended to facilitate the surveillance or censorship of citizens' online

[34] *See* Leviathan Security Group *Quantifying The Costs of Forced Localisation* (2015)https://static1.squarespace.com/static/556340ece4b0869396f21099/t/559dad76e4b0899d97726a8b/1436396918881/Quantifying+the+Cost+of+Forced+Localization.pdf

activities and violate individuals' human rights including freedom of speech, expression, access to information, and privacy and due process rights.

The legislation's interaction with other laws is unclear

Statutory reform

Whilst we support the goal of reducing the harm associated with online harassment, the legislation relates to defamation. This change would interact with a set of evolving and complex new requirements.

The Council of Attorneys-General defamation law reform process is still underway. Stage 1 of this defamation law reform has been implemented in some, but not all, states and territories. Stage 2 of the reform process, which specifically deals with intermediary liability, is ongoing. The Stage 2 discussions have involved a careful and thoughtful consideration of the complex issues associated with intermediary liability.

The "anti-trolling" legislation raises new concepts which have not been part of the Stage 2 discussion. The legislation - and the state and territory defamation laws - should align on important concepts such as formal requirements for complaints, defences available to intermediaries, and a common approach to obtaining details of anonymous users.

Case law

Additionally, the law is still being developed through key cases currently in the courts. The case of *Voller*, which was explicitly identified by the Government as a key driver for the legislation, is yet to be finalised, and importantly the applicability of defences, including the innocent dissemination defence for page owners, has yet to be determined in that case.

Furthermore, an appeal to the High Court is pending in the case of *Defteros*, another case, in which, like *Voller*, the Court will consider issues of online publication for defamation.

Alignment with defamation laws

Currently, there is a significant disconnect between state and territory laws and the legislation. For example, under the Stage 1 reforms, a complainant cannot commence

proceedings without first issuing a concerns notice which identifies the online location of defamatory content (e.g. by uniform resource locator- URL) and states the "serious harm" caused by the defamation. In contrast, a complainant under the legislation need not provide these details when making a complaint. There should be consistency between the complaints scheme under the legislation and the concerns notice process under state and territory laws.

It is also possible that two separate liability regimes could apply to the same piece of content, which could lead to inconsistent results. For example:

- If a user located in Australia and a user located in New Zealand both post the same defamatory comment on a social media page, the social media provider would be deemed a publisher of the comment posted by the user located in Australia (and would not have the benefit of the innocent dissemination defence) but would not be deemed a publisher of the comment posted by the user located in New Zealand (and would still retain the benefit of the innocent dissemination defence). It is not clear why a different liability regime should apply to a social media provider, depending on where the poster is located. Especially because defamation law is focussed on the place of publication and not the location of the author or originator of the content.
- If a Page owner posts defamatory material in the form of a news article on their own website, the legislation will not apply to that material and existing defamation laws will apply. However, if a Page owner posts the same defamatory material in the form of a news article on their social media page, the new law will apply and the social media provider will be deemed to be the publisher of news article (and will not be entitled to rely on other defences including the innocent dissemination defence). The rationale for this difference in approach is unclear and could lead to inconsistency.

Existing online safety regulations and complaint scheme

The Australian Government has very recently updated Australian online safety law, via the Online Safety Act passed in July 2021. For any Australians seeking redress against online harassment, the complaint scheme administered by the Office of the eSafety Commissioner provides sufficient redress.

The recently enacted Online Safety Act has only just taken effect in January 2022. And yet, this legislation is duplicative and apparently seeking to advance the same policy objectives that underpinned the Online Safety Act.

The most significant element of the Online Safety Act will be a new scheme that allows eSafety to order the takedown of online material that is bullying or harassing an Australian adult (previously, eSafety's takedown scheme for bullying or harassment content was limited to children). eSafety has indicated that this scheme will be available to public figures, and will include online content that includes claims of criminal conduct and character attacks.[35] Depending on the specific formulation of a piece of online content, this could in effect set a **lower** threshold than defamation law (for example, because it applies to all claims of criminal conduct and does not provide a defence of truth).

Two other powers of eSafety are important to note here, given the elements of the legislation:

1. eSafety also has the ability to issue end user notices directly to individuals who are perpetrators of content captured by their takedown schemes, including bullying or harassment of an Australian adult.
2. eSafety already has data disclosure powers that can compel a social media platform to provide available contact information of a user on their services. This has only recently begun to be used in relation to Meta's services.

These mechanisms are more suited to address the harms associated with online harassment than defamation law, which is focussed on compensation for damage to reputation.

[35] *See* Adult Cyber Abuse Scheme Regulatory Guidance December 2021 eSC RG 3 https://www.esafety.gov.au/sites/default/files/2021-12/ACA%20Scheme%20Regulatory%20Guidance%20%20FINAL.pdf

The defence sets an impossible bar to meet

The legislation removes a social media provider's ability to rely on the innocent dissemination defence in respect of material posted in Australia. This is a significant departure from Australia's existing defamation laws, as well as the position in other common law jurisdictions, such as the United Kingdom. The removal of the defence means that social media providers will be liable for content as soon as it is posted and before they are even aware of its existence. This is a fundamental shift in the defamation landscape and puts social media providers in the same position as primary publishers, despite the fact that they do not have the same level of editorial control over, or knowledge of the content.

While the legislation does seek to introduce a new "complaints scheme" defence for social media providers, the requirements of that new defence set an impossible bar for social media providers to meet. Coupled with the loss of the innocent dissemination defence, social media providers are likely to be left without a defence in many circumstances, despite their best efforts to remove harmful content from their platforms and to comply with the requirements of a complaints scheme.

Definition of relevant contact details

The definition of "relevant contact details" is likely to have a substantial impact on Australian users of social media services. "Relevant contact details" are defined to mean the name of the person, a phone number that *can be used to contact the person* and an email address that *can be used to contact the person*.

In order to rely on the new defence, a social media provider must disclose the "relevant contact details" of the user who posted the material. In order to obtain the benefit of the new defence in all circumstances, social media service providers will be incentivised to:

- mandate the collection of full name, phone number **and** email address from all users in Australia;
- verify that the name provided by such users is their real name or the name by which they are usually known (e.g. by verifying ID documents);
- verify that all users in Australia can be contacted using such information on a regular basis (e.g. by sending intrusive verification text messages and emails to users which require active confirmation by those users);
- limit the ability of all users in Australia to delete such information (without replacing it with new, verified information);
- retain the information of all users in Australia who have deleted their accounts in case future complaints are made against those users; and

- restrict access to services for those users in Australia who do not provide such information or who fail to verify such information (for any reason).

The above processes could have a disproportionately negative impact on certain groups of users, including those without access to a mobile phone or those without access to ID documentation. They could also stifle freedom of expression as users may no longer feel comfortable expressing themselves if there is a risk that their identity and contact details could be disclosed to another user. This could have a particularly chilling effect on whistleblowers, survivors of sexual assault, victims of domestic violence and other users who could be put at risk if their identity was disclosed. It may also reduce the extent to which users of social media engage in political speech, including by criticising elected officials.

Authenticity is the cornerstone of Meta's community. We believe authenticity helps create a community where people are accountable to each other, and to Meta, in meaningful ways. But we want to allow for the range of diverse ways that identity is expressed across our global community, while also preventing impersonation and identity misrepresentation. Authentication should be tailored to the specific risk seeking to be mitigated and proportionate in impacting the users implicated in the risk.

There are also significant issues with this definition from a practical perspective. There is simply no way that a social media service provider could guarantee that a user will always be contactable using the information they have provided to the provider. Even if a provider uses best efforts to verify such information, at any point in time, a user could change their phone number, be locked out of their email account or simply refuse to respond to contact (especially from a complainant). In addition, dedicated bad actors will always find ways to circumvent verification systems, including by using VPNs to mask their location (and therefore evade verification checks focussed on Australian users), by using other people's phone numbers or email addresses to respond to verification checks or by purchasing 'burner' phones or creating 'burner' emails and then disposing of them.

In addition, the requirement to provide "relevant contact details" seems disproportionate to the purpose of the definition. The Explanatory Memorandum states that the purpose of the definition is to "is to enable defamation proceedings to be commenced, including by way of substituted service if authorised by a court".[36] However, it is possible to effect substituted service for the purpose of defamation proceedings using only a phone

[36] Page 7.

number **or** an email address.[37] Meta requires that users provide either a phone number or email address in order to sign-up to Facebook or Instagram. There is no clear rationale for the requirement to provide both of these contact points. In addition, courts have also granted orders for substituted service via social media.[38] This obviates the need for social media service providers to disclose any contact details in some cases (such as where the court is satisfied that service is not otherwise practicable and the proposed method of service is likely to bring the documents to the notice of the defendant).

No defence if content is removed

In Meta's experience, a complainant typically wishes to have defamatory content removed from a social media service as quickly as possible. However, under the legislation, removal of content will not provide a defence to a social media provider. The only defence available to the provider will be the new complaints scheme defence. This could have the consequence of exacerbating harm to a complainant as the social media service provider would have little incentive to remove the content while the complaints scheme process is playing out. As a social media provider has no editorial control over the content posted by its users, it should have a defence in circumstances where it expeditiously removes defamatory content on notice of such content.

No defence if court does not make order

A social media service provider could be unfairly left without a defence where a complainant seeks an end-user information disclosure order and the court has not yet issued or refuses to issue the order. There may be a number of reasons for this, including because disclosure is likely to present a risk to the poster's safety[39] or because the complainant does meet the criteria for such an order e.g. because the complainant is not an Australian person (even though a non-Australian person is entitled to bring proceedings against the social media provider) or because the complainant is able to ascertain the relevant contact details of the poster.[40] A social media provider should be entitled to a defence in these circumstances.

---

[37] Rule 10.24 Federal Court Rules 2011; see also, *Nettle v Cruse* [2021] FCA 935 (11 August 2021) at para [8]. Wigney J made an order deeming service of a claim by an email address known to have been used by the defendant. That order was then served on the defendant via text message.
[38] Wakim v Criniti [2016] NSWSC 1723; A & K Collins Investments Pty Ltd v Keto Pumps S A R L [2020] WASC 231 at [8]; Re RH; Ex parte RH by next friend CH [2020] WASC 13 from [82] to [89]; Queensland Building & Construction Commission v van Uden [2021] QDC 103 from [16] to [20].
[39] See cl 19(3).
[40] See cl 19(1)(b) and cl 19(1)(c)(i) and 19(2)(e).

A court may also limit the scope of an order to the disclosure of "relevant contact details" or "country location data" (not both) or to something less than "relevant contact details" (e.g. name and email address only). In these circumstances, compliance with the order would not give the social media service provider a defence as the defence only applies where the provider has disclosed both "relevant contact details" *and* "country location data" (regardless of the scope of the court order)[41]. A court may, for example, only order the disclosure of country location data because the complainant is able to ascertain the relevant contact details of the poster in another way. This does not seem reasonable, given that the social media provider is left without a defence due to factors outside of its control.

No defence if complainant is vexatious

The legislation states that a social media service provider is not required to take action in response to a complaint or request to disclose relevant contact details if it reasonably believes that the complaint or the request does not genuinely relate to the potential institution by the complainant of a defamation proceeding against the poster in relation to the material[42]. The Explanatory Memorandum indicates that, despite this, a social media service provider "would not be able to have access to the defence"[43]. This does not seem reasonable as it means that a provider will be forced to engage with a vexatious complainant (including those who may have been declared vexatious by the courts) throughout the complaints scheme in order to preserve its defence, even though the complainant has no genuine intention of commencing defamation proceedings against the poster. This is particularly concerning in the case of complainants who may seek to abuse the complaints process to find out the identity and contact details of other users for other purposes. A social media provider should have a defence if it reasonably believes that the complaint or the request does not genuinely relate to the potential institution by the complainant of a defamation proceeding against the poster in relation to the material.

Complaints scheme requirements

The requirements of the complaints scheme are extremely onerous and will make it difficult for social media providers to operationalise the scheme. In particular:

- **Requirements for complaint:** The legislation does not prescribe any requirements for a complaint under the scheme. This is in direct contrast to the requirements for

[41] Cl 16(2)(d)(ii)
[42] Cl 17(1)(i).
[43] Page 19.

a "concerns notice" under state and territory laws,[44] which must: (1) be in writing; (2) specify the location of the defamatory matter (e.g. webpage address); (3) state the asserted defamatory imputations; (4) state the serious harm caused or likely to be caused to the complainant's reputation by publication. A complainant cannot (without the court's leave) commence defamation proceedings without first giving a concerns notice. The legislation should align the requirements for a complaint under the complaints scheme with the requirements for a concerns notice in order to avoid duplicative processes and ensure that a social media provider has adequate information in order to respond to a complaint in a timely manner.

- **Response times:** The 72-hour time frames proposed in the complaints scheme process are onerous, particularly where a complaint may be vague, lack critical information (such as a URL for the material) or refer to a large volume of material. This is exacerbated by the fact that, as discussed above, the legislation does prescribe requirements for the complaint. These timeframes are also strict. If the provider is even just one hour late in responding to a complaint, it may lose the benefit of the defence. There may be a good reason for the delay, including where the provider is working to locate the content, despite inadequate identifying information from the complainant, or where there are complex freedom of speech issues at play. The complainant and the provider may also agree to an extension of time in some circumstances. It does not seem proportionate that the provider should lose the defence in these circumstances. A more balanced approach should be taken, such as requiring complaints to receive a response within a reasonable timeframe or "without undue delay".
- **Channel for complaints:** The legislation does not specify when a complaint is 'made' to a social media provider for the purposes of determining when the 72 hour turnaround period commences. The lack of clarity on when a complaint is 'made', coupled with the fact that there is no specified channel for receiving such complaints, will make it extremely difficult for providers to operationalise the 72 hour turnaround period. Social media service providers should be able to specify a single point or channel for complaints (e.g. online form or email address) in order to effectively manage them given the short turnaround times. Without this, it is likely that complaints will be made through a variety of different (and incorrect) channels, which will be difficult to manage and will jeopardise the provider's ability to respond within the specified timeframes. While the Government has introduced a right for it to prescribe legislative rules in relation to how a provider is required to communicate with a complainant, there is no limitation on how a complainant is required to communicate with a provider.

[44] See Defamation Act 2005 (NSW), incorporating Stage 1 amendments.

Other concerns

Broad scope of the law

The scope of content covered by the Bill is very broad. While the Exposure Draft only applied to "comments" made on a page, the Bill now applies to all "material" posted on a page. "Material" is defined to include any material such as text, speech, music, visual images or any other type of data.[45] This definition would capture all content posted on social media, including photos, videos, podcasts, ads, news articles, media broadcasts, etc. This change dramatically increases the scope of the Bill, particularly when coupled with the broad definition of "page".

In addition, it represents a significant departure from the type of content considered by the High Court of Australia in *Voller.* In *Voller*, the High Court of Australia considered comments which had been posted by third parties on Facebook Pages administered by certain news companies. Given that one of the main drivers behind the Bill is to address the issues raised by the decision in Voller, this expansion goes beyond what is necessary.

We recommend that the Government revert to use of the term "comment". In addition, we recommend that "comment" be defined to reflect the way in which the term is used in a social media context (i.e. a reactive written response to content posted on a platform by another user).

In addition, the Bill has the potential to apply to a broad range of services, beyond those that would be considered as social media services under the Online Safety Act. Unlike the Online Safety Act, the Bill does not include a separate definition of "relevant electronic service", which captures services such as email, instant messaging, SMS, MMS and interactive online games. Without a separate definition for "relevant electronic services", the definition of "social media services" under the Bill could be interpreted very broadly to capture services that would be considered to be "relevant electronic services" under the Online Safety Act. This interpretation does not appear to be consistent with the Government's intention. We note that the Government has the power to include or exclude certain services from the definition through legislative rules. However, this ambiguity will create substantial uncertainty for businesses, especially given the significant product and operational changes required to comply with the Bill. The definition of "social media service" should therefore expressly exclude "relevant electronic services".

[45] Section 5 of the *Online Safety Act 2021*.

The draft legislation also deals with liability in a defamation proceeding which "relates to" material. The term "relates to" is quite vague and could potentially capture proceedings that do not directly concern the relevant material.

Collection and verification of identity information may raise privacy concerns

As noted above, the requirement to hold "relevant contact details" will incentivise foreign social media social providers, such as US and Chinese based companies, to collect and regularly verify contact details of their users and to restrict access for those users who refuse to provide or verify these details. The legislation therefore runs counter to fundamental data minimisation concepts by incentivising the collection of additional data about Australians which is not necessary to provide services.

While the definition of "relevant contact details" currently requires the collection of name, email address and phone number, there is scope for the definition to be expanded to include other categories of information by legislative rules. From a privacy perspective, this is particularly concerning as the definition could be further expanded to include more intrusive information such as date of birth or physical address for a large group of Australians participating online.

The legislation also incentivises practices which are inconsistent with the principles underpinning some of the proposals put forward in the Discussion Paper for the Privacy Act Review. As mentioned above, the legislation incentivises providers to refuse to provide services to users who have not provided or verified their contact details in order to protect themselves from liability for content posted by those users. This appears to be inconsistent with the Government's aim of increasing consumers' ongoing control over their personal information by introducing a "right to object" to the collection, use or disclosure of personal information.[46]

Finally, the legislation may incentivise some providers to disclose the contact details of a user without that user's consent in order to protect themselves against liability. This is because, once a complainant has made a request for contact details under the complaints scheme, the provider is only entitled to rely on the defence if such details are actually disclosed by the provider (regardless of whether or not the user consents to such disclosure). This may encourage some providers to balance the privacy risk associated

[46] *See* Privacy Act Review Discussion Paper, p114 https://consultations.ag.gov.au/rights-and-protections/privacy-act-review-discussion-paper/user_uploads/privacy-act-review-discussion-paper.pdf

with disclosing personal information without consent, against the risk of liability for defamation under the legislation. This is particularly concerning in circumstances where there may be strong reasons for a user to withhold consent to disclosure, such as a risk to their safety.