

# STATUTORY REVIEW OF THE ONLINE SAFETY ACT 2021

The Centre of the Public Square at Per Capita

June 2024

The Centre of the Public Square (CPS) at Per Capita welcomes the opportunity to provide this submission to the Department of Infrastructure, Transport, Regional Development, Communications and the Arts (Online Safety, Media and Platforms Division) as part of their *Statutory Review of the Online Safety Act 2021*.

Per Capita is an independent think tank, dedicated to fighting inequality in Australia. We work to build a new vision for Australia, based on fairness, shared prosperity, and social justice. CPS works to build better models of citizen collaboration and strengthen civil society by imagining new methodologies and alternate technologies to anchor this public space.

## Key points

- Australia has developed online safety legislation that has left us with strong laws addressing content online, but weak laws addressing and overseeing the risk-producing systems themselves.
- Social media platform providers should be responsible for the impact their business decisions have on what is essentially public space - the burden of responsibility for online harms should fall on them and not solely on individual users.
- Australian lawmakers should take a risk management approach in online safety legislation - digging beneath the content layer to target the risk producing systems themselves.
- Laws should be amended to impose a **statutory duty of care on platforms** to ensure, so far as is reasonably practicable, that the users of their service, and that people who may be affected by the service and are not users of that service, are not appreciably harmed as a result of its operation or use.
- This duty of care should be broad; singular; and enforceable by the regulator.

## PER CAPITA SUBMISSION

- Platforms should be required to take proactive steps to manage, minimise, and mitigate reasonably foreseeable harms that may arise from the use of their services. This action would better future proof our online safety system as providers would be required to constantly assess new and emerging potential harms.
- Implementing and enforcing a duty of care will require additional resourcing and staff which will come with additional cost to the public. It is appropriate to recover costs for eSafety regulatory activities from industry. This is provided for under UK and EU regulations.
- Strong transparency measures will need to be implemented to ensure appropriate steps are being taken to identify and mitigate risks and comply with the duty of care. Australian laws makers should incorporate aspects that are similar to those in the *EU Digital Services Act* that deal with transparency requirements into our own legislation. These are best practise and large global companies are already working under these rules in some parts of the world.
- Large social media platforms have been a democratising force and arguably make up our modern town square. Laws need to be carefully designed to ensure freedom of expression online is not unreasonable burdened while also considering other fundamental human rights.
- Allowing certain researchers to access data and interrogate the operations of platforms increases transparency and accountability, aids government in understanding how to appropriately regulate the online sphere, and provides users with the ability to make an informed choice about whether or not to participate on platforms.
- Large social media companies have shown reluctance to act in the best interest of their users on numerous occasions. It has become clear that governments must regulate for transparency, accountability and compliance with human rights by these companies.
- Additional requirements on social media platforms should be proportionate to their size and user base in Australia.

## Introduction

Australia has developed an online safety regulatory framework that has left us with strong content focused laws. This involves complaints and content-based removal notice schemes to support those who have experienced online harm. However, it lacks provisions to address and oversee the risk-producing systems themselves.

Globally, there is an emerging trend in digital service regulation that seeks to compel greater accountability, transparency and compliance with human rights from large social media companies and digital service providers. These address the *impact* that providers have on what is essentially public space and focus on what their responsibilities and obligations to the public should be as stewards of the modern public square. Australia should take a similar approach.

## Rise of the digital agora

Over the last 30 years, rapid technological development and digitalisation has changed everything about how we live our lives. COVID-19 lockdowns and closures only increased our reliance on the internet as workplaces, schools and social activities moved online. Social media platforms have engendered a revolution in communication. Here, people can connect with each other, unconstrained by physical location. Social media platforms, like Facebook, X, and YouTube, provide cheap and accessible avenues for communication. On these platforms, everyday citizens are no longer mere recipients of information; they now act in many roles: as producers, curators, and distributors.

With 20.8 million social media users in Australia (78.3% of the total population), and 5.04 billion social media users around the world,<sup>1</sup> social media platforms have carved out a central role in our social, cultural and political lives. They often function as the 'central arena in which political change is initiated, negotiated, organized and communicated'.<sup>2</sup> There is little doubt that platforms have become the stages for essential political debate in Australia and across the globe, evidenced by the many social and political movements which simply could not have achieved their global reach, nor affected real social, cultural and political change, without social media platforms acting as conduits.<sup>3</sup>

---

<sup>1</sup> Simon Kemp, 'Digital 2024: Australia', *DataReportal* (Web Page, 21 February 2024) <https://datareportal.com/reports/digital-2024-australia>; Simon Kemp, 'Digital 2024: Global Overview Report', *DataReportal* (Web Page, 31 January 2024) <https://datareportal.com/reports/digital-2024-global-overview-report>.

<sup>2</sup> Maik Fielitz and Karolin Schwarz, 'Hate Not Found?! Deplatforming The Far Right and Its Consequences' (Research Report, Institut für Demokratie und Zivilgesellschaft, December 2020) 9.

<sup>3</sup> For example, #metoo, occupy, black lives matter etc.

Social media platforms can be a democratising force, but it is equally true that they facilitate societal and individual harms. The same mechanisms that have allowed for the widespread impact of positive social justice movements have also enabled the dissemination of hateful, extremist, and harmful abuse online. This can, and does, spill out of the digital sphere and into the real world. Instances of cyberbullying are increasing online, leading to numerous types of harm experienced by individuals and entire sections of our community. More and more, Australian's are reporting negative online experiences. The eSafety 2022 *Adult's Negative Online Experiences Survey* found that 75% of respondents had at least one negative online experience in the previous 12 months, up from 58% in the 2019 survey.<sup>4</sup> 28% of respondents said these negative online experiences had a moderate to extreme impact on their mental or emotional wellbeing, and 16% said it had a moderate to extreme impact on their physical wellbeing.<sup>5</sup> Children are similarly impacted. In a 2021 survey of children aged 12-17, 44% reported having at least one negative online experience in the six months to September 2020,<sup>6</sup> and Headspace's 2020 *National Youth Mental Health Survey* found that over half of young Australians believe their mental health is declining, with 42% attributing social media as the primary factor for this decline.<sup>7</sup>

## Platforms moderate and governments regulate

In the last decade, governments around the world have begun abandoning their previously *hands-off* approach to social media regulation and have moved to regulate the online sphere with greater readiness and enthusiasm than ever before. Since 2016, over one hundred inquiries and reviews into platform regulation have commenced globally.<sup>8</sup> Australia has held several reviews and inquiries concerning online safety,<sup>9</sup> and our Federal Parliament has enacted laws aimed at protecting users' online safety and well-being.<sup>10</sup>

The path Australia has taken in developing its online safety regime has been heavily content-based. Reset.Tech Australia, a research and policy organisation specialising in digital risk and

<sup>4</sup> eSafety Commissioner, Australian Government, 'Online Safety in Australia - Adult's Experiences Online' (Infographic, 6 February 2023) <https://www.esafety.gov.au/research/australians-negative-online-experiences-2022/infographic-adults-experiences-online>.

<sup>5</sup> Ibid.

<sup>6</sup> eSafety Commissioner, Australian Government, *The Digital Lives of Aussie Teens* (Report, February 2021) <https://www.esafety.gov.au/sites/default/files/2021-02/The%20digital%20lives%20of%20Aussie%20teens.pdf?v=1717475602873>.

<sup>7</sup> These findings indicate a notable rise compared to the 2018 survey, in which only 37% identified social media as the catalyst for mental health decline: 'Young People Cite Social Media as Main Reason for Worsening Mental Health', Headspace (News Post, 9 June 2022) <https://headspace.org.au/our-organisation/media-releases/young-people-cite-social-media-as-main-reason-for-worsening-mental-health/>.

<sup>8</sup> Dwayne Winseck and Manuel Puppis, 'Platform Regulation Inquiries, Reviews and Proceedings Worldwide' (unpublished, February 2023) [https://docs.google.com/document/d/1AZdh9sFCGfTQEROQjo5fYeiY\\_gezdf\\_11B8mQFsuMfs/edit](https://docs.google.com/document/d/1AZdh9sFCGfTQEROQjo5fYeiY_gezdf_11B8mQFsuMfs/edit).

<sup>9</sup> See, eg, Lynette Briggs, *Report of the Statutory Review of the Enhancing Online Safety Act 2015 and the Review of Schedules 5 and 7 to the Broadcasting Services Act 1992 (Online Content Scheme)* (Report, October 2018); Australian Competition and Consumer Commission, *Digital Platforms Inquiry* (Final Report, June 2019); Senate Standing Committee on Environment and Communications, Parliament of Australia, *Online Safety Bill 2021 [Provisions] and Online Safety (Transitional Provisions and Consequential Amendments) Bill 2021 [Provisions]* (Final Report, March 2021); House of Representatives Select Committee on Social Media and Online Safety, Parliament of Australia, *Social Media and Online Safety* (Report, March 2022); Senate Standing Committee on Legal and Constitutional Affairs, Parliament of Australia, *Social Media (Anti-Trolling) Bill 2022 [Provisions]* (Report, March 2022).

<sup>10</sup> Explanatory Memorandum, Online Safety Bill 2021 (Cth) 1. See also Explanatory Memorandum, Criminal Code Amendment (Sharing of Abhorrent Violent Material) Bill 2019 (Cth) 9 [28].

online harms, asserts that this has ‘left us with strong regulations regarding content but very weak regulatory oversight on the risk-producing systems themselves’.<sup>11</sup>

The focus of our laws has primarily been on defining prohibited content on platforms, providing for a complaints and content removal scheme, and enforcing compliance through civil and criminal penalties on platform providers. This carries negative ramifications for freedom of expression and robust political discourse online. As private companies their primary imperative is to generate profit. Thus, penalising providers for the content on their platforms incentivises them to err on the side of caution, over-censoring expression to protect their private interests.<sup>12</sup>

Purely content focused regulation may have been appropriate in the traditional media context, but the volume of content posted and interacted with by users on social media is astronomical, dwarfing the output of traditional media. In just one 2022 minute, over 347 thousand Twitter users shared a tweet, 66 thousand Instagram users shared a photo, 500 hours of video was uploaded by YouTube users, and Facebook users shared 1.7million pieces of content globally.<sup>13</sup> Regulation of social media based on traditional media is not suited to the online sphere. Content moderators and regulators cannot keep up with the volume of content posted every day. As one piece of harmful content is addressed, many more appear in its place.

All platforms moderate content. Not doing so would be unsustainable for their businesses. As researcher Tarleton Gillespie asserts: ‘moderation is, in many ways, *the* commodity that platforms offer’.<sup>14</sup> They do this to ensure users keep participating on their platforms, to reassure advertisers that the platform is conducive to their commercial interests, and to demonstrate to lawmakers that further regulation is unnecessary.<sup>15</sup> User safety is not their main concern – users are not their most influential stakeholders. The business of social media companies is not in public advocacy, it is in advertising, and advertisers are their largest stakeholder. In 2022, YouTube’s Worldwide advertising revenues reached \$29.2 billion (USD);<sup>16</sup> and between 2010 and 2022 an average of 94% of Meta’s yearly revenue was generated by advertising.<sup>17</sup> Decisions made by YouTube in 2017, to implement tighter rules on inappropriate

<sup>11</sup> Reset.Tech Australia, ‘A Duty of Care in Australia’s Online Safety Act’ (Policy Briefing Document, April 2024)

<https://au.reset.tech/uploads/Duty-of-Care-Report-Reset.Tech.pdf>.

<sup>12</sup> See, eg Evelyn Douek, ‘Australia’s “Abhorrent Violent Material” Law: Shouting “Nerd Harder” and Drowning Out Speech’ (2020) 94(1) *Australian Law Journal* 41, 47; Human Rights Council, *Report of the Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression*, 32<sup>nd</sup> sess, UN Doc A/HRC/32/38 (11 May 2016).

<sup>13</sup> Data Never Sleeps’, *Domo* (Web Page, 2023) <https://www.domo.com/data-never-sleeps>.

<sup>14</sup> Tarleton Gillespie, *Custodians of the Internet: Platforms, Content Moderation, and the Hidden Decisions That Shape Social Media* (Yale University Press, 2018) 13 (Emphasis in original).

<sup>15</sup> *Ibid* 47.

<sup>16</sup> Laura Ceci, ‘Youtube’s Advertising Revenues as Percentage of Google’s Global Revenues From 2017 to 2022’, *Statista* (Web Page, 6 February 2023) <https://www.statista.com/statistics/289658/youtube-global-net-advertising-revenues/>. This revenue is split between YouTube and YouTube creators (55% and 45% respectively).

<sup>17</sup> Stacy Dixon, ‘Meta: Annual Revenue 2009-2022, by Segment’, *Statista* (Web Page, 19 April 2023)

<https://www.statista.com/statistics/267031/Facebooks-annual-revenue-by-segment/>.

content alongside other major changes to their moderation policies, did not occur with the aim of reducing online harm for the benefit of users.<sup>18</sup> Rather, it was to retain unhappy advertisers who had begun abandoning the platform in what has been described as YouTube’s ‘Adpocalypse’.<sup>19</sup>

As Facebook whistleblower Frances Haugen told the House Select Committee on Social Media and Online Safety in 2022: ‘Platforms cannot be trusted to act in the public interest. They are often...fully aware of the harms caused by their products and services, and yet choose to ignore these in favour of growth and profit’.<sup>20</sup>

Perhaps a simple response to the harms people face on social media platforms is just to stop using these services, but it is not that simple. The parent organisations of YouTube and Facebook (Google and Meta) are among the ten most valuable corporations in the world.<sup>21</sup> These digital platforms wield extraordinary global dominance and monopoly over the digital media sector, allowing them to have control over both what we communicate and how we communicate.<sup>22</sup> In the public sphere, removing yourself from social media means removing yourself from the ‘stages for political debate’,<sup>23</sup> as well as an increasingly major source for accessing news. In the private sphere these companies are just as ubiquitous. Each day people send 100 billion direct messages and make 150 million calls over Facebook Messenger alone.<sup>24</sup>

Facebook CEO Mark Zuckerberg has said that:

[i]n a lot of ways Facebook is more like a government than a traditional company...[w]e have this large community of people, and more than other technology companies we’re really setting policies.<sup>25</sup>

Facebook is not a country. The X board is not a government. It is true that providers have carved out a role as pseudo-governors of the digital public square, but they have taken few steps to act in the best interests of the people who populate it. They should be responsible for the space they have created and own, much like property owners in the physical world. Real

<sup>18</sup> Neerav Srivastava, ‘Indie Law for YouTubers: YouTube and the Legality of Demonetisation’ (2021) 42(2) *Adelaide Law Review* 503, 504-5.

<sup>19</sup> Ibid. See also Sangeet Kumar, ‘The Algorithmic Dance: YouTube’s Adpocalypse and the Gatekeeping of Cultural Content on Digital Platforms’ (2019) 8(2) *Internet Policy Review* <https://policyreview.info/articles/analysis/algorithmic-dance-youtubes-adpocalypse-and-gatekeeping-cultural-content-digital>.

<sup>20</sup> Evidence to House Select Committee on Social Media and Online Safety, Parliament of Australia, Canberra, 3 February 2022, 1 (Frances Haugen).

<sup>21</sup> Matthew Johnston, ‘Biggest Companies in the World by Market Cap’, *Investopedia* (Web Page, 24 September 2022) <<https://www.investopedia.com/biggest-companies-in-the-world-by-market-cap-5212784#toc-6-tesla-inc-tsla?>>.

<sup>22</sup> Terry Flew, Fiona Martin and Nicolas Suzor, ‘Internet Regulation as Media Policy: Rethinking the Question of Digital Communication Platform Governance’ (2019) 10(1) *Journal of Digital Media & Policy* 33, 34.

<sup>23</sup> David Bromell, *Regulating Free Speech in a Digital Age: Hate, Harm and the Limits of Censorship* (Springer, 2022) 32.

<sup>24</sup> Nicola Bleu, ‘27 Latest Facebook Messenger Statistics (2023 Edition)’, *Blogging Wizard* (Web Page, 1 January 2023) <<https://bloggingwizard.com/facebook-messenger-statistics/>>.

<sup>25</sup> Mark Zuckerberg quoted in David Kirkpatrick, *The Facebook Effect: The inside Story of the Company That Is Connecting the World* (Simon & Schuster, 2011) 254.

governments can strongly influence how platform providers operate their services and what decisions they make. They can implement laws that compel greater accountability, transparency and compliance with human rights. In relation to online harm, there will always need to be some focus on content and assisting people who have experienced and encountered harmful content, but our laws should go further, digging beneath the content layer to target the risk producing systems themselves.

Everything that happens on social media is the result of some business decision. They make decisions about their terms of service, user guidelines, their use of software (including AI and automatic filtering), moderation processes, and resourcing.

Implementing a single and broad statutory duty of care in the *Online Safety Act 2021* (Cth) will incentivise social media companies to consider the impact their decisions have on users and to take proactive steps to managed, minimise and mitigate online harms that may be enabled by them - flipping the responsibility of maintaining online safety from users to service provider.

### A statutory duty of care

Prior to the drafting and passage of the UK's *Online Safety Act 2023* (OSA) academics Lorna Woods and William Perrin developed a public policy proposal for the implementation of a statutory duty of care and regulator to reduce online harm experienced by internet users in the UK.<sup>26</sup> The focus of their proposal was not on content directly, but rather it was on service design. Woods and Perrin explained that:

[t]his duty of care does not equate to direct liability for the publication of content or for its distribution, but rather constitutes a recognition that the consequences of design and business choices should be taken into account and, where practicable, dealt with.<sup>27</sup>

The UK OSA implemented multiple duties of care, as opposed to a single broad duty of care proposed by Woods and Perrin. It includes general duties on all providers of regulated user-to-user services and search services in relation to illegal content risk assessments, illegal content, content reporting, complaints procedures, freedom of expression and privacy, and record keeping and review; and additional duties are applicable to services that meet certain thresholds based on size, functionality, reach and risk potential.<sup>28</sup> There are some issues arising from these multiple duties as opposed to a single broad duty of care. It requires distinguishing between different types

---

<sup>26</sup> Lorna Woods and William Perrin, *Online Harm Reduction – A Statutory Duty of Care and Regulator* (Report, Carnegie UK, April 2019) [https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=4003986](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4003986).

<sup>27</sup> Lorna Woods and William Perrin, 'Obliging Platforms to Accept a Duty of Care' in Martin Moore and Damian Tambini (eds), *Regulating Big Tech* (Oxford University Press, 2021) 93, 95.

<sup>28</sup> *Online Safety Act 2023* (UK) pt 2-3.



of content and assigning duties to each. This somewhat pulls the regulations back towards a *content-first* over a *system-first* approach. Reset.Tech Australia explains that:

[a] systemic approach acknowledges that systems are developed and business decisions are made about them before they are populated with content. Requiring platforms to think about risk assessing their systems only after they are 'populated' with particular sorts of content reduces the broader efficacy of the approach.<sup>29</sup>

Australia should adopt a single statutory duty of care, as opposed to multiple duties like those implemented in the UK's OSA. This should be a statutory duty of care on platforms to ensure, so far as is reasonably practicable, that the users of their services are free from harm arising from its operation or use, and that people who may be affected by the service and are not users of that service are not appreciably harmed as a result of its operation or use.<sup>30</sup>

This duty is broad, and it should be. There are a variety of harms people can be exposed to online. Morgan Scheuerman et al categories four types of harm through their work categorising online harm: physical, emotional, relational, and financial.<sup>31</sup> Placing a broad duty of care will require platforms to take proactive steps to mitigate reasonably foreseeable harms by identifying risks. They will need to consider risks as they apply to all potential users and people affected by the operation of their services including users of different ages, genders and cultural backgrounds. A broad statutory duty of care is also relatively future-proof. Platforms will necessarily need to consider and plan accordingly for new and emerging risks. This is particularly important considering the speed in which the online environment continues to evolve bringing new challenges including the development and prevalence of artificial intelligence.

A statutory duty of care would need to be enforceable and enforced by a regulator able to prosecute for breaches of the duty. It should be backed with penalties depending on severity. There is already a complaints mechanism under the Australian *Online Safety Act*,<sup>32</sup> which could be expanded to provide for complaints and investigations relating to breaches of the duty of care.

Implementing and enforcing a duty of care will require additional resourcing and staff which will come with additional cost. This cost should not be purely borne by the public. Other jurisdictions provide for online services in certain circumstances to pay a fee to cover the costs of regulation to

---

<sup>29</sup> Reset.Tech Australia, 'A Duty of Care in Australia's Online Safety Act' (Policy Briefing Document, April 2024) 8.

<sup>30</sup> Text is based on proposed text in Lorna Woods and William Perrin, *Online Harm Reduction – A Statutory Duty of Care and Regulator* (Report, Carnegie UK, April 2019).

<sup>31</sup> Morgan Klaus Scheuerman et al, 'A Framework of Severity for Harmful Content Online' (2021) 5 (CSCW2) *Proceedings of the ACM on Human-Computer Interaction* <https://dl.acm.org/doi/pdf/10.1145/3479512>.

<sup>32</sup> *Online Safety Act 2021* (Cth) pt 3.



keep the online environment safe for users.<sup>33</sup> Social media companies are some of the most valuable companies in the world. Meta and Twitter alone had a \$128 billion (USD) combined market capitalization as of March 2024.<sup>34</sup> It is appropriate to recover costs for the eSafety's regulatory activities from industry.

Strong transparency measures will need to be implemented to ensure appropriate steps are being taken to identify and mitigate risks and comply with the duty of care. The EU *Digital Services Act* (DSA) provides a number of transparency measures that Australia should adopt. For example, *Very Large Online Platforms* (VLOPs) (online intermediaries with more than forty-five million monthly) must undertake risk assessments yearly on systemic risks; put in place reasonable, proportionate, and effective mitigation measures tailored to systemic risks; and undertake an independent audit to assess compliance with their obligations under the DSA.<sup>35</sup> They are required to report to both a Digital Services Coordinator (DSC) and the public on the results of their risk assessments and specific mitigation measures put in place.<sup>36</sup> Additionally, they must publish transparency reports every six months on a range of matters related to how they operate and moderate their services including information on their content moderation teams including their qualifications and linguistic expertise.<sup>37</sup>

The DSA also provides for data access and scrutiny. VLOPs must provide access to data that is necessary to monitor and assess compliance to a DSC and to vetted researchers who are investigating systemic risks.<sup>38</sup> It is in the public interest for government and the public to understand the extent of online risks. This can better be achieved by allowing certain researchers to access data and interrogate the operations of platforms.

## Conclusion

The *Online Safety Act 2021* (Cth) commenced in January 2022 and was introduced to 'create a modern, fit for purpose regulatory framework that builds on the strengths of the existing legislative scheme for online safety'.<sup>39</sup> Arguably this framework is already out of date. The Australian Government should amend the *Online Safety Act* with an additional focus on systems. That is, moving beyond platform accountability for the content they carry and towards accountability for the design and governance of their services and harm producing systems. In order to achieve this the Australian Government should implement a broad statutory duty of care on social media

<sup>33</sup> *Online Safety Act 2023* (UK) pt 2-3; *Regulation (EU) 2022/2065 of the European Parliament and of the Council of 19 October 2022 on a Single Market For Digital Services and amending Directive 2000/31/EC (Digital Services Act)* [2022] OJ L 277/1, art 43 (*EU Digital Services Act*).

<sup>34</sup> Stacy Dixon, 'Social Media Companies Mau and Market Cap at IPO 2024', *Statista* (Web Page, 3 April 2024)

<https://www.statista.com/statistics/1459282/mau-market-cap-at-ipo-social-media-companies/>.

<sup>35</sup> *EU Digital Services Act* (n 33) arts 34 – 37.

<sup>36</sup> *Ibid* art 42

<sup>37</sup> *Ibid* art 42.

<sup>38</sup> *Ibid* art 40.

<sup>39</sup> Explanatory Memorandum, *Online Safety Bill 2021* (Cth) 1.

platforms; require platforms to regularly identify, assess, and mitigate risks; and ensure compliance with additional enforcement powers and transparency requirements.