

Tech reform and rebuild

Submission to the Inquiry into Social Media and
Online Safety

Jordan Guiao

Research Fellow, Centre for Responsible Technology

December 2021

About The Australia Institute

The Australia Institute is an independent public policy think tank based in Canberra. It is funded by donations from philanthropic trusts and individuals and commissioned research. We barrack for ideas, not political parties or candidates. Since its launch in 1994, the Institute has carried out highly influential research on a broad range of economic, social and environmental issues.

About the Centre for Responsible Technology

The Australia Institute established the Centre for Responsible Technology to give people greater influence over the way technology is rapidly changing our world. The Centre will collaborate with academics, activists, civil society and businesses to shape policy and practice around network technology by raising public awareness about the broader impacts and implications of data-driven change and advocating policies that promote the common good.

Our philosophy

As we begin the 21st century, new dilemmas confront our society and our planet. Unprecedented levels of consumption co-exist with extreme poverty. Through new technology we are more connected than we have ever been, yet civic engagement is declining. Environmental neglect continues despite heightened ecological awareness. A better balance is urgently needed.

The Australia Institute's directors, staff and supporters represent a broad range of views and priorities. What unites us is a belief that through a combination of research and creativity we can promote new solutions and ways of thinking.

Our purpose - 'Research that matters'

The Institute publishes research that contributes to a more just, sustainable and peaceful society. Our goal is to gather, interpret and communicate evidence in order to both diagnose the problems we face and propose new solutions to tackle them.

The Institute is wholly independent and not affiliated with any other organisation. Donations to its Research Fund are tax deductible for the donor. Anyone wishing to donate can do so via the website at <https://www.tai.org.au> or by calling the Institute on 02 6130 0530. Our secure and user-friendly website allows donors to make either one-off or regular monthly donations and we encourage everyone who can to donate in this way as it assists our research in the most significant manner.

Level 1, Endeavour House, 1 Franklin St
Canberra, ACT 2601
Tel: (02) 61300530
Email: mail@tai.org.au
Website: www.tai.org.au
ISSN: 1836-9014

Summary

Australia, like many other countries around the world has increasingly been trying to grapple with the many harms and issues the public can encounter online.

The largest online and social media platforms, like Facebook/Meta and YouTube have been shown to promote harmful disinformation, have algorithms which prize harmful content, and create damaging issues for individuals, groups, and institutions.

A range of regulatory interventions have been proposed and enacted over the last 12/24 months. Unsurprisingly, those that work best are mandatory initiatives which force action from the online platforms, backed by legislation and penalties, and propose broad reviews, while the less effective ones are those which continue to place minimal burden on platforms, or facilitate a self-regulation regime with narrow scope.

The Australia Institute's Centre for Responsible Technology welcomes the opportunity to submit to the Inquiry into Social Media and Online Safety, and propose the following recommendations:

- **Pursue systemic regulatory interventions** which account for the complexity and breadth of online platform harms (e.g. broad system-wide reviews like the Privacy Act and Adtech inquiry works best, whereas narrow reviews like defamation in social media channels are limited)
- **Progress the ACCC digital platform inquiry recommendations** which to date have generated the most impactful outcomes and considers the broad scope of online platforms
- **Build and develop alternative digital infrastructure** to provide Australians with alternative, safe and transparent digital platforms so they can migrate away from private and harmful platforms. A publicly funded digital platform coded with ethical algorithms and transparent frameworks would provide a safe digital alternative.

Introduction

The House Select Committee on Social Media and Online Safety was established by a resolution of appointment that passed the House of Representatives on December 1, 2021. The Committee is conducting an inquiry into social media and online safety.

The Australia Institute's Centre for Responsible Technology welcomes the opportunity to submit to this inquiry. This submission addresses the following specific points in the inquiry's terms of reference:

- the range of harms that may be faced by Australians on social media and other online platforms, including harmful content or harmful conduct
- evidence of the potential impacts of online harms on the mental health and wellbeing of Australians
- evidence of the extent to which algorithms used by social media platforms permit, increase or reduce online harms to Australians
- the effectiveness, take-up and impact of industry measures, including safety controls, protections and settings, to keep Australians, particularly children, safe online
- the transparency and accountability required of social media platforms and online technology companies regarding online harms experienced by their Australian users
- actions being pursued by the Government to keep Australians safe online

For the purposes of this submission, we will also focus on two of the largest and most damaging online platforms – Facebook/Meta, and YouTube.

Harms online and on social media

The evidence for online harms caused by platforms like Facebook and YouTube are now well-documented.

On September/October 2021, a series of damning internal documents on Facebook, later revealed to be leaked by whistleblower Francis Haugen were published. The documents detailed several issues with Facebook, clearly demonstrating harms online:

- Internal Facebook research showed that its mobile app Instagram was harmful to young users, particularly among teenage girls.¹
- Facebook was aware of the hate speech and calls for violence against minority ethnic groups in India, but did little to intervene.²
- Facebook knew about harmful conspiracy theories circulating in the lead up to the United States 2020 elections.³
- For years, Facebook's algorithms promoted posts which provoked angry reactions, as they generated more engagement than those which generated positive or neutral reactions. Further, staff attempted to intervene on this issue, but was overruled by CEO Mark Zuckerberg, who feared any intervention would lead to less engagement.⁴
- Staff continuously expressed concerns over the harms Facebook is causing but upper management did little to resolve these.⁵
- There was evidence of the platform being used to facilitate human trafficking.⁶
- Facebook's internal moderation resources to combat online harms and harmful content overly relied on AI rather than human moderators who can navigate the subtlety, context and complexity of different content.⁷

¹ Gayle (2021), *Facebook aware of Instagram's harmful effect on teenage girls, leak reveals*, <https://www.theguardian.com/technology/2021/sep/14/facebook-aware-instagram-harmful-effect-teenage-girls-leak-reveals>

² Frenkel & Alba (2021), *In India, Facebook grapples with an amplified version of its problems*, <https://www.nytimes.com/2021/10/23/technology/facebook-india-misinformation.html>

³ Timberg et. al. (2021), *Inside Facebook, Jan.6 violence fueled anger, regret over missed warning signs*, <https://www.washingtonpost.com/technology/2021/10/22/jan-6-capitol-riot-facebook/>

⁴ Merill & Oermus (2021), *Five points for anger, one for a 'like': How Facebook's formula fostered rage and misinformation*, <https://www.washingtonpost.com/technology/2021/10/26/facebook-angry-emoji-algorithm/>

⁵ Hendel (2021), *'This is NOT normal': Facebook employees vent their anguish*, <https://www.politico.com/news/2021/10/25/facebook-employees-message-anguish-517012>

⁶ Duffy (2021), *Facebook has known it has a human trafficking problem for years. It still hasn't fully fixed it*, <https://edition.cnn.com/2021/10/25/tech/facebook-instagram-app-store-ban-human-trafficking/index.html>

⁷ Culliford & Heath (2021), *Facebook knew about, failed to police, abusive content globally – documents*, <https://www.reuters.com/technology/facebook-knew-about-failed-police-abusive-content-globally-documents-2021-10-25/>

This litany of issues is only one recent instance of the revelations against Facebook/Meta which has had sustained problems with online harms in its platforms that continue unresolved.

Google-owned YouTube has similar systemic issues with online harms, which continue unresolved:

- YouTube has similar issues with content moderation which allows disinformation to go unchecked, including around the pandemic and political campaigns during elections.⁸
- Despite clear content policies, YouTube's algorithm recommends videos which violates these policies, including content with misinformation, violence, hate speech and scams.⁹
- There are questions about the role of YouTube's algorithms in radicalising individuals, and the lack of transparency and clarity on the algorithm's workings leaves this question unanswered, with YouTube unwilling to provide access.¹⁰
- During a Senate hearing YouTube was asked to defend why they only perform "tweaks and minor changes" when it comes to protecting against harmful content, particularly for younger users and children.¹¹

Facebook/Meta and YouTube are two of the largest online platforms in the world, used by millions of Australians and their platforms continue to facilitate harms online, with their self-regulatory efforts maintain an ineffective regime to protect against these harms.

⁸ Nguyen & Scurato (2021), *Facebook and YouTube's refusal to moderate misinformation in global languages harms communities of color*, <https://prismreports.org/2021/11/02/facebook-and-youtubes-refusal-to-moderate-misinformation-in-global-languages-harms-communities-of-color/>

⁹ Stokel-Walker (2021), *YouTube's algorithm recommends videos that violate its own policies*, <https://www.newscientist.com/article/2283354-youtubes-algorithm-recommends-videos-that-violate-its-own-policies/>

¹⁰ Camargo (2020), *YouTube's algorithms might radicalise people – but the real problem is we've no idea how they work*, <https://theconversation.com/youtubes-algorithms-might-radicalise-people-but-the-real-problem-is-weve-no-idea-how-they-work-129955>

¹¹ Gordon (2021), *Senator: YouTube, TikTok, Snap offering only small changes*, <https://techxplare.com/news/2021-10-youtube-tiktok-snap-execs-senators.html>

Current proposals

The platforms' efforts in self-regulation to date have been ineffective in protecting against harms online, as detailed in the previous section.

Government intervention has shown some impact, with regulatory efforts by the Australian government forcing platforms into action.

However, there has been varying degrees of success with these regulatory efforts and many are currently still in development:

- The Voluntary Code for Dis- and Misinformation administered by the Australian Communications and Media Authority (ACMA) was an industry code developed by the Digital Industry Group Inc. (DIGI) – a lobby group funded by the Big Tech companies including Meta and Google.¹² The Code has been criticised as “laughable”¹³ and “woefully inadequate”¹⁴ as it provides minimal intervention with the urgent problems around disinformation. The policies include business-as-usual tactics from the online platforms, and the process relies on the public actively reporting on breaches of the Code. The Code has no material penalties for breaches and the independent oversight board tactic chosen to oversee the process has been proven to be ineffective overseas.¹⁵ Overall the Code does little to address significant disinformation challenges online.
- The Online Safety Act updates have provided some significant updates in combating online harms via the eSafety Commissioner, including increased penalties against breaches and increased capabilities.¹⁶ There are industry codes currently in development to further clarify content criteria. The process has however, been criticised for lack of public and community consultation and due process, with the legislative updates seemingly rushed through¹⁷ and the current industry code development largely consulting with industry groups, and not the wider community. The eSafety office must balance its interventions and not overreach themselves

¹² DIGI (2021), *DIGI website*, <https://digi.org.au/disinformation-code/>

¹³ Smith & Ward (2021), *Tech giants' 'laughable' disinformation solution slammed*, <https://www.afr.com/technology/tech-giants-laughable-disinformation-solution-slammed-20211010-p58ys8>

¹⁴ Sadler (2021), *Big Tech misinformation efforts slammed as 'woefully inadequate'*, <https://www.innovationaus.com/big-tech-misinformation-efforts-slammed-as-woefully-inadequate/>

¹⁵ McKay (2021), *Facebook says it just can't keep up with Oversight Board's recommendations*, <https://www.gizmodo.com.au/2021/11/facebook-says-it-just-cant-keep-up-with-oversight-boards-recommendations/>

¹⁶ eSafety Commission (2021), *Online Safety Act 2021 Fact sheet*, <https://www.esafety.gov.au/sites/default/files/2021-07/Online%20Safety%20Act%20-%20Fact%20sheet.pdf>

¹⁷ Sadler (2021), *Why the rush? Online Safety Bill still not passed*, <https://www.innovationaus.com/why-the-rush-online-safety-bill-still-not-passed/>

which critics are worried about.¹⁸ Their initiatives would benefit from sunset clauses, further community consultation on takedown criteria and protections like appeals processes.

- Defamation on social media/anti-trolling laws looks at the ability for individuals to pursue defamation action against comments made about them on social media platforms, and the ability to unmask “trolls” online.¹⁹ While a crackdown on trolling activity as well as the coordinated weaponisation of content through bots and anonymous trolls are welcome, many criticise the proposed Bill as too narrowly focused, difficult to enforce and exposing vulnerable users online for which anonymity is a protection.²⁰ This proposal should also be wary of not being a self-serving tool for public figures who want to minimise legitimate criticism against them.
- The Privacy Act review and the draft Online Privacy Bill shows early promise with wide-ranging proposals giving individuals more control over their data, more penalties for privacy breaches and increased capabilities for the Privacy Commissioner.²¹ Given the data economy powers the largest online platforms, interventions against data abuses and online privacy could be a significant way of curbing online harms via the platforms. The Privacy Act proposals appear to have effective countermeasures as the proposals consider the complexity and breath of the online landscape, with a range of actions that consider individuals but also groups and institutions. Given the tendency for online platforms to defy conventional definitions and industry divisions, a wholesale and holistic review of the online landscape and the data economy is an effective way of ensuring online platforms cannot use the excuse of complexity to avoid regulatory measures.

¹⁸ Stardust (2021), *A new online safety bill could allow censorship of anyone who engages with sexual content on the internet*, <https://theconversation.com/a-new-online-safety-bill-could-allow-censorship-of-anyone-who-engages-with-sexual-content-on-the-internet-154739>

¹⁹ Karp & Taylor (2021), *What are the Coalition’s proposed anti-troll social media laws and who do they benefit?* <https://www.theguardian.com/media/2021/dec/01/who-really-benefits-from-the-coalitions-proposed-anti-troll-social-media-laws>

²⁰ Floreani (2021), *Online anonymity is really important, actually*, <https://overland.org.au/2021/10/online-anonymity-is-really-important-actually/>

²¹ Australian Attorney-General’s Department (2021), *Privacy Act Review Discussion Paper*, https://consultations.ag.gov.au/rights-and-protections/privacy-act-review-discussion-paper/user_uploads/privacy-act-review-discussion-paper.pdf

Recommendations/Conclusion

Online harms are clearly evidenced in online platforms. The largest companies like Meta and YouTube continue to have issues in their platforms. Their self-regulation efforts are not working and government intervention is necessary. However, intervention works best when they are mandatory and consider the complexity of the digital landscape. The Centre for Responsible Technology therefore recommends that the Government:

- **Pursue systemic regulatory interventions** which account for the complexity and breadth of online platform harms (e.g. broad system-wide reviews like the Privacy Act and Adtech inquiry works best, whereas narrow reviews like defamation in social media channels are limited)
- **Progress the ACCC digital platform inquiry recommendations** which to date have generated the most impactful outcomes and considers the broad scope of online platforms
- **Build and develop alternative digital infrastructure** to provide Australians with alternative, safe and transparent digital platforms so they can migrate away from private and harmful platforms. A publicly funded digital platform coded with ethical algorithms and transparent frameworks would provide a safe digital alternative.