



Response to the Finance and Public Administration Reference Committee's Inquiry into Administration of the referendum into an Aboriginal and Torres Strait Islander Voice

Executive summary

Australia's regulatory framework is neither comprehensive nor rigorous enough to address the threats posed by electoral mis and disinformation, including threats likely to emerge in the upcoming Voice referendum. Broader regulatory requirements that hold platforms accountable for the promotion of mis and disinformation, coupled with requirements for transparency to enable effective independent oversight, are urgently needed.

These requirements could take the shape of the EU's *Digital Services Act*, or through requiring specific 'duties of care' from platforms to users in Australia. We note that there is both movement in this broad direction, and potential political will to address this issue through an update and review of the *Online Safety Act*.

Understanding of the effectiveness of platforms' responses to mis and disinformation is key to enabling robust regulatory responses. During the referendum, Reset.Tech Australia will be running a monitoring and analysis project with the following core activities:

- Reviews of platform's existing terms of service, highlighting content moderation gaps or failures in complaint mechanisms that may affect electoral mis and disinformation
- Evaluation of the capacities made available to bad actors by platforms to target individuals, including vulnerable individuals, with mis and disinformation
- Evaluation of the 'take down' responses from platforms, where mis and disinformation is reported
- Review of the advertising and advertisers associated with funding mis and disinformation content
- Analysis of the algorithmic amplification of mis and disinformation on platforms, including engagement and growth metrics for disinformation content or actors

This is to enable an informed consideration of the regulatory models that might work best to secure Australia's information architecture during electoral cycles. Active consultation and escalation channels will also ensure our project permits timely responses to threats as they emerge.

We especially welcome feedback on the following:

- What an appropriate 'ad hoc' threat escalation channel could look like for any emerging threats we identify during this referendum, and;
- Who the expected threat actors are, and the key narratives to specifically monitor, from the perspective of ensuring the integrity of the referendum process.

Contents

| | |
|---|---|
| Executive summary | 2 |
| Contents | 3 |
| 1. About Reset.Tech Australia & this submission | 1 |
| 2. About the Susan McKinnon Foundation | 1 |
| 3. Monitoring platform responses to mis and dis information, to inform regulatory action | 2 |
| A. Australia's current policy approach to address electoral mis and dis information | 2 |
| B. Evaluating platform responses to mis and disinformation in the Voice referendum | 5 |
| C. Longer term policy solutions for securing Australia's information architecture during electoral cycles | 6 |
| 4. Conclusions | 7 |
| 5. Appendix | 8 |

1. About Reset.Tech Australia & this submission

[Reset.Tech Australia](#) is an independent, non-partisan policy initiative and research organisation. We are the Australian affiliate of [Reset.Tech](#), a global initiative working to counter digital threats to democracy. Reset.Tech has extensive experience in monitoring electoral mis and dis information with a focus on identifying areas for regulatory intervention.

This submission has been prepared in response to the Finance and Public Administration Reference Committee's Inquiry into Administration of the referendum into an Aboriginal and Torres Strait Islander Voice. It focuses on point B of the inquiry's terms of reference, around the 'detection, mitigation, and obstruction of potential dissemination of misinformation and disinformation', specifically focussing on social media.

We will outline Reset.Tech Australia's Voice mis and dis information monitoring plan for the Voice referendum, and how this could inform long term policy development.

2. About the Susan McKinnon Foundation

We are grateful to the Susan McKinnon Foundation for supporting our work on mis and disinformation in the Voice referendum. Information on the Susan McKinnon Foundation's mission and objectives is as follows:

Susan McKinnon Foundation is a nonpartisan, not-for-profit organisation that works to help Australia achieve a more fit-for-purpose political, policy and service delivery system. The Foundation was established by Grant Rule and Dr Sophie Oh with the aim of making a lasting difference to Australia by helping to enhance the capability and effectiveness of our democratic institutions and government. Misinformation can have a corrosive effect on our democratic processes and institutions by misleading voters, suppressing voter turnout, and eroding trust in democratic institutions. Our primary objective is to promote the public interest and not to support a particular agenda. By supporting initiatives that work to counter the impacts of misinformation SMF seeks to provide Australians with the opportunity to make decisions based on accurate and reliable information.

3. Monitoring platform responses to mis and dis information, to inform regulatory action

Mis and disinformation is rife in the Australian information architecture, including electoral mis and dis information. For example;

- A QUT study examined around 54,000 Twitter accounts during and after the 2019 Australian Federal Election (looking at over 1 million tweets). It found that 13% of accounts were 'very likely' to be bots, with the majority originating from New York.¹ This is estimated to be **more than double the rate of bot accounts in the US presidential election**. These can have big impacts: research into the US election by ANU indicated that the average bot was 2.5 times more influential than the average human, measured by success at attracting exposure via retweets.²
- Chinese Australians have faced misinformation in the past, often in what appear to be **coordinated disinformation campaigns**.³ Social media platforms, such as WeChat, Weibo and Douyin have been found to serve targeted misinformation to Chinese language speakers in Australia. In 2019, WeChat in particular was a site of much political campaigning in Mandarin which included mis & disinformation.⁴

Despite the risks, Australia's current regulatory framework does not have a strong nor comprehensive approach to electoral mis and disinformation. It lies outside the scope of the *Online Safety Act* and the eSafety Commissioner's remit, and beyond the reach of the Australian Electoral Commission.

A. Australia's current policy approach to address electoral mis and dis information

Australia's policy response to electoral mis and dis information is limited, and like many of our digital platform policies, is industry-drafted and co-regulatory. It is largely left to the Disinformation and Misinformation Code of Practice ('DIGI Code'). Industry drafted, co-regulator models suffer from two significant constraints; industry-led drafting creates sub-standard levels of protection,⁵ and, the inevitably voluntary nature of these efforts create

¹ See study quoted in Felicity Caldwell (2019) 'Bots stormed Twitter in their thousands during the federal election' *SMH*
www.smh.com.au/politics/federal/bots-stormed-twitter-in-their-thousands-during-the-federal-election-20190719-p528s0.html

² Sherryn Groch (2018) 'Twitter bots more influential than people in US election: research' *SMH*
www.smh.com.au/national/twitter-bots-more-influential-than-people-in-us-election-research-20180913

³ Kirsty Lawson (2020) 'WeChat the channel for China disinformation campaigns' *Canberra Times*
<https://www.canberratimes.com.au/story/6802076/the-social-messaging-system-helping-spread-chinese-disinformation-campaigns/>

⁴ Kirsty Lawson (2020) 'WeChat the channel for China disinformation campaigns' *Canberra Times*
<https://www.canberratimes.com.au/story/6802076/the-social-messaging-system-helping-spread-chinese-disinformation-campaigns/>

⁵ Reset Tech (2022) *How outdated approaches to regulation harm children*
<https://au.reset.tech/news/how-outdated-approaches-to-regulation-harm-children-and-young-people-and-why-australia-urgently-needs-to-pivot/>

coverage issues.⁶ The problems are systemic to the model, and not isolated to Dis and Misinformation Code of Practice. For example, the Online Safety Codes drafted through this process are poised to be rejected by the eSafety Commissioner.⁷

The systematic weaknesses of industry drafted Codes are not limited to Australia. The DIGI Code closely imitates the European Union's first attempt at online content and safety policy – the *Code of Practice on Disinformation* (2018). In the European experience, policy decision makers soon discovered the Code suffered from transparency and measurability constraints, as below:

*At present, it remains difficult to precisely assess the timeliness, comprehensiveness and impact of the platforms' actions, as the Commission and public authorities are still very much reliant on the willingness of platforms to share information and data. The lack of access to data ... (along with) the absence of meaningful KPIs to assess the effectiveness of platform's policies to counter the phenomenon, is a fundamental shortcoming of the current Code.*⁸

The 2018 Code was eventually replaced by a revised version in 2022, which has been further galvanised and strengthened by provision in the EU's *Digital Services Act*.

A similar policy trajectory is already visible in Australia. The Government has announced an intention to reinforce the DIGI Code and give the ACMA more powers, specifically to register an enforceable industry code and to set standards, should industry self-regulation measures prove insufficient in addressing the threat posed by misinformation and disinformation.⁹

Likewise, the Government has announced that in recognition of the gaps in protection created by the Online Safety Act, they are committed to reviewing it earlier than the January 2025 requirement, to ensure our 'world-leading online safety framework remain(s) fit for the changing online environment'.¹⁰ We would suggest that the goal should be to improve Australia's legislative framework, to ensure the same levels of protection against mis and dis information provided across Europe.

⁶ For example BitChute, Odyssey and Telegram are not signatories despite being available in Australia and known vectors of disinformation and misinformation. See: Adobe, Apple, Google, Meta, Microsoft, Redbubble, TikTok and Twitter. See ACMA (2022) *Australian Code of Practice for Disinformation and Misinformation*
<https://www.acma.gov.au/online-misinformation#:~:text=you%20have%20concerns.-,Australian%20Code%20of%20Practice%20for%20Disinformation%20and%20Misinformation,%2C%20Redbubble%2C%20TikTok%20and%20Twitter.>

⁷ Brandon How (2023) 'Concerns raised over draft online safety codes several times' *InnovationAus*
<https://www.innovationaus.com/concerns-raised-over-draft-online-safety-codes-several-times/>

⁸ European Commission (2020) 'Staff Working Document: Assessment of the Code of Practice on Disinformation - Achievements and areas for further improvement'. Found at:
<https://digital-strategy.ec.europa.eu/en/library/assessment-code-practice-disinformation-achievements-and-areas-further-improvement>

⁹ Minister for Communications (2023) *New ACMA powers to combat harmful online misinformation and disinformation*
<https://minister.infrastructure.gov.au/rowland/media-release/new-acma-powers-combat-harmful-online-misinformation-and-disinformation>

¹⁰ Australian Government (2023) *Australian Government response to the House of Representatives Select Committee on Social Media and Online Safety report*
<https://www.infrastructure.gov.au/sites/default/files/documents/australian-gov-response-to-house-of-reps-select-committee-on-social-media-and-online-safety-report-march2023.pdf>

The boxes below demonstrate that both the EU and UK approaches centralise transparency from platforms, and for regulators to hold them accountable for the responses to mis and disinformation hosted and promoted on their platforms.

Policy approach to mis and dis information in the EU

In Europe, the Digital Service Act places transparency and accountability obligations on platforms who disseminate mis and disinformation. It creates specific obligations to:

- Conduct a systematic risk assessment of their platform at least once a year (for Very Large Online Platforms). Very large online platforms will have to mitigate against these risks, or face action from regulators.
- Requirements around transparency and to allow 'vetted' independent researchers to access data. This should additionally allow independent verification of platform's risk assessments.
- Enable user appeals, through an internal complaints mechanism and an additional out-of-court settlement process.

Policy approach to mis and dis information in the UK

In the UK, the Online Safety Bill¹¹ proposes a series of 'duties of care' from platforms to users that may address mis and disinformation. For example, it provides for:

- Duties around hosting and promoting illegal content,
- Duties to provide users with more control over the content they are shown,
- Duties to uphold their terms of service, including content moderation practices that may address mis and dis information
- Duties about complaints processes
- Duties to provide transparency reports about particular types of content

It also proposes establishing a regulatory advisory committee on mis and disinformation that is empowered to advise OFCOM, the regulator, around how to best exercise their functions.

Understanding what an 'adequate' response to mis and disinformation looks like will be critical for regulators in the UK (to understand how they might be failing in their duties as described above) and the EU (to understand the risk assessments created by platforms, evaluating their mitigation processes, allowing researchers to compare responses and enabling users to take informed complaints). In the EU context, Reset.Tech is working to develop a set of metrics around mis and disinformation to help regulators evaluate these risk assessments.

Building on this, Reset.Tech Australia is undertaking an assessment of platforms' responses to mis and disinformation during the Voice referendum in Australia. The aim is not to 'admire the problem' and describe the nature and types of mis or disinformation spread during the Voice referendum. Nor is it to recreate a fact checking service. Rather, we are aiming to comprehensively evaluate how platforms respond to or promote this mis and disinformation, including developing key metrics that platforms could be held accountable to. This is a more systemically focussed task, aiming to demonstrate how Australian regulators could evaluate platform's responses and ultimately create a regulatory system that ensures transparency and delivers accountability.

¹¹ UK Government (2022) Online Safety Bill
<https://bills.parliament.uk/publications/49376/documents/2822>

B. Evaluating platform responses to mis and disinformation in the Voice referendum

We are planning on a comprehensive monitoring of platform responses to mis and disinformation in the Voice referendum. The table below sets out our focus areas.

| Activity | Rationale |
|--|--|
| Reviews of platforms' existing terms of service, highlighting content moderation gaps or failures in complaint mechanisms. | Regulators may consider holding platforms to account for achieving their own content and moderation 'standards' as laid out in their terms of service. It is also necessary to identify where platforms' terms of service may not adequately address Australian specific mis and disinformation threats, and where new regulation may be needed to step in, in the long term. In the short term, these reviews may be helpful for platforms to identify where they can and should step up throughout the referendum. |
| Evaluation of the capacities made available to bad actors by platforms to target individuals, including vulnerable individuals, with mis and disinformation. | Evaluate the extent platforms prevent or facilitate bad actors to interfere with Australian electoral processes, such as through acts of coordinated inauthentic behaviour. |
| Evaluation of the 'take down' responses from platforms when mis and disinformation is reported. | Evaluate the effectiveness of platforms' existing take down and notice practices, which is the main pathway to recourse under the current <i>Online Safety Act</i> . ¹² |
| Review of the advertising and advertisers associated with funding mis and disinformation content. | Mis and disinformation is not only a content problem, it is a product of, and responsive to, various market forces. The review will place the business model driving mis and disinformation into full focus, and the role platforms play in creating a marketplace for this content. |
| Analysis of the algorithmic amplification of mis and disinformation on platforms, including engagement and growth metrics for disinformation content or actors. | Assess and understand the role of platforms and their systems in promoting mis and disinformation content. |

¹² Note, the Online Safety Act currently acts as a 'backstop' for certain types of content. Electoral mis and disinformation is not covered under the current scheme.

While the focus of our work is on the need for and shape of future regulation, this process will also involve identifying threats to Australia's information architecture as they emerge. Ideally, we would like to be able to action these insights as they arise through threat escalation channels. This requires identifying and establishing actors capable of receiving and responding to emerging threats. **We would welcome engagement from this Committee to help us identify and develop appropriate escalation channels.**

Undertaking this work relies on accurately identifying key threat actors and threat narratives. **We welcome this Committee's advice identifying Indigenous leaders, experts and community organisations to engage in our consultation. We also welcome feedback on pertinent threat actors to capture in our source lists.**

C. Longer term policy solutions for securing Australia's information architecture during electoral cycles

The need for 'ad hoc' escalation channels highlights the core, systemic difficulty facing Australia's electoral landscape. There is no permanent regulatory mechanism available to seek recourse for and to tackle identified risks. While Australia lacks a clear policy framework for these sorts of threats, there are policy models for addressing them internationally. In Europe, Reset.Tech affiliates have been developing metrics to support regulators in this task. We aim to draw upon these processes and the shape metrics to the Australian context and the Voice referendum. **We would be delighted to connect the Committee with our experts in the EU to provide evidence or discuss this further if this is helpful.**

Metrics informing mis and disinformation regulation in the EU

Reset.Tech affiliates in the EU have been monitoring mis and disinformation in a range of global elections, including elections in the UK, the US, Kenya, Germany, France and Brazil, with a view to understanding the sorts of metrics that can be derived to enable regulatory action.

With the *Digital Services Act* taking effect across the EU, regulators are able to demand proportionate and proactive mitigation from social media platforms to reduce the risks of mis and dis information. Reset.Tech have developed a set of metrics for the European context, including the below. We would be happy to share the full list with the Committee.

- Average engagement with disinformation vs genuine content
- Average growth rate for disinformation pages/actors vs genuine pages/actors
- Non follower engagement rates (on YouTube and Twitter)
- Content moderation indicator (Response and notice reaction rates)
- Average toxicity score of comments, by actor or #

While the regulatory mechanisms for actions are currently not in place in Australia, we note that the Government has expressed a commitment to reviewing the scope of the *Online Safety Act* to understand potential gaps,¹³ which may include electoral mis and dis information. Reset.Tech Australia welcomes this commitment to review any gaps in protections created by the currently online safety frameworks, and recommends that this should include protection from mis and disinformation. This is in keeping with our broader calls for a more comprehensive regulatory framework that addresses the societal risks created by digital platforms, as well as the individual risks.¹⁴ **If it is helpful to this Committee, we are happy to discuss alternate international proposals for digital platform regulations including mis and dis information.** A summary is available in the Appendix.

4. Conclusions

We anticipate Australia's regulatory framework is not comprehensive enough to address the threats posed by electoral mis and disinformation, including within the Voice referendum. Ultimately, we believe that broader requirements are needed to hold platforms accountable for the promotion of mis and disinformation, coupled with requirements for transparency facilitating effective independent oversight. Future legislation may take the shape of the EU's *Digital Services Act*, or through requiring specific duties of care from platforms to users in Australia. Either approach requires an accurate baseline of platforms' current responses to mis and disinformation. During this referendum, Reset.Tech Australia will both identify threats as they emerge and monitor platforms' responses to mis and disinformation, culminating in evidence-led recommendations for future legislation and regulatory models.

¹³ Australian Government (2023) *Australian Government response to the House of Representatives Select Committee on Social Media and Online Safety report*
<https://www.infrastructure.gov.au/sites/default/files/documents/australian-gov-response-to-house-of-rep-s-select-committee-on-social-media-and-online-safety-report-march2023.pdf>

¹⁴ See, for example, Reset.Tech (2022) *The Future of Digital Regulation in Australia*
<https://au.reset.tech/uploads/the-future-of-digital-regulations-in-australia.pdf>

5. Appendix

Figure One: Comparative approaches to types addressing harms through regulation

| | EU | GERMANY | UK | IRELAND | CANADA | AUSTRALIA |
|---|---|--|---|--|--|---|
| Key legislation addressing harms | Digital Services Act (in force) | NetzDG, and others (in force) | Online Safety Bill (in draft) | Online Safety & Media Regulation (recently passed) | Online safety proposals (currently being redrafted) | Online Safety Act (in force) |
| Definition of Harm, Individual, Community or Societal | No set definition, the focus is on harms that violate rights. This will include societal harms, and community harm through hate speech | Based on existing criminal law. This includes Individual and some community harms through hate speech | Individual (Content having an adverse physical or psychological response on adults or children) | Individual (Illegal content, individually intimidating or threatening content, eating disorder, self harm & suicide content) | Individual (aligned to existing definitions of hate speech) Societal (damage to societal cohesion, vulnerable groups) | Individual (content that is "offensive" to adults or children, content that is refused classification etc) |
| Focus on systems or Takedown | Systems + Takedown | Takedown | Systems + Takedown | Systems + Takedown | Takedown | Takedown (+ potentially some systems through BOSE and co-regulatory Codes) |
| Content In Scope | Illegal + indirectly, legal Disinfo included indirectly Hate speech indirectly included | Illegal Disinfo out of scope Hate speech in scope | Illegal + legal List of harms to be added later but unclear whether disinfo & hate speech is in scope (could be in scope where content is harmful to adults) | Illegal + legal Disinfo out of scope Individual hate speech content could be in scope, where it intimidates, threatens, humiliates or persecutes | Illegal Disinfo out of scope Hate speech in scope | Illegal + legal Disinfo out of scope Individual hate speech content could be in scope, where it causes offence to an individual or would be considered menacing, harassing or offensive |
| Services In Scope | Intermediary services | Social media | Services which host | Broad range of | Social media | Social media services, |

| | | | | | | |
|---------------------------|---|-------------|---|---|--|---|
| | e.g. ISPs and online platforms Private messaging out of scope | | or facilitate UGC, apart from news media outlets. Private messaging in scope. | platforms and services inc press publications which enable UGC Private messaging in for criminal content | Private messaging out of scope | Relevant electronic service and ISPs (Tight definition of "social media") |
| Powers Of Regulator | Fines Information gathering powers Algorithmic audit mandatory | Fines | Fines Information gathering powers Language seems to allow algorithmic inspection | Fines Information gathering powers. No algorithmic audit | Information gathering powers Inspection powers No algorithmic audit | Fines Offers public facing complaint mechanisms, Investigation, Audit (not algorithmic) |
| Independence Of Regulator | Independent as well as EC oversight of large platforms | Independent | Independent however OSB keeps provisions for political agenda setting | Independent Creates Online Safety Commissioners | Independent Creates Digital Safety Commissioner and Digital Recourse Council of Canada, | Independent |
| Transparency | Six monthly transparency reports (publicly published) Data access for pre-vetted researchers | | Annual transparency reports No data sharing provisions | Periodic transparency reporting | Transparency reporting inc data on takedown volumes and processes. | Transparency reporting |