CYBER SECURITY
COOPERATIVE
RESEARCH
CENTRE

SUBMISSION:

Communications Legislation Amendment (Combatting Misinformation and Disinformation) Bill 2023 – Exposure Draft

cybersecuritycrc.org.au

To whom it may concern,

Submission: *Communications Legislation Amendment (Combatting Misinformation and Disinformation) Bill 2023* – Exposure Draft

I am pleased to submit the Cyber Security Cooperative Research Centre's (CSCRC) response to the Department of Infrastructure, Transport, Regional Development, Communications and the Arts' consultation on the *Communications Legislation Amendment (Combatting Misinformation and Disinformation) Bill 2023* (the Bill).

There is no doubt that steps need to be taken to stem the flow of harmful misinformation and disinformation in Australia. Misinformation and disinformation – which are largely cyber-enabled via digital platforms – have the potential to erode our democracy, social systems and economy, with a deleterious impact on our way of life. However, while the proposed Bill's intended functions are theoretically sound, it has the potential to result in serious unintended consequences like, for example, impacting freedom of speech. Furthermore, if the Bill is enacted, appropriate resourcing must be provided to ensure the Australian Communications and Media Authority (ACMA) can fulfil its duties, as regulating digital platforms at such scale will be significantly difficult.

About the CSCRC

The CSCRC is dedicated to fostering the next generation of Australian cyber security talent, developing innovative projects to strengthen our nation's cyber security capabilities. We build effective collaborations between industry, government and researchers, creating real-world solutions for pressing cyber-related problems.

By identifying, funding and supporting research projects that build Australia's cyber security capacity, strengthen the cyber security of Australian businesses and address policy and legislative issues across the cyber spectrum, the CSCRC is a key player in the nation's cyber ecosystem.

We look forward to answering any queries about this submission and welcome the opportunity to participate in any future consultation.

Yours Sincerely,

Rachael Falk

CEO, Cyber Security Cooperative Research Centre

[redacted]

Introduction

Australian public trust in the ability to receive information online is at an all time low, with the *Edelman Trust Barometer 2023* finding that just 25% of Australians trust social media platforms.[1]

While there is a clear need to counter the spread of mis- and disinformation, this must be balanced with the unintended consequences that may arise from regulation, the most serious of which could be a chilling effect on freedom of expression and public discourse. While the CSCRC understands this is not the intention of the *Communications Legislation (Combatting Misinformation and Disinformation) Bill 2023* (the Bill), the broad definitions that have been applied and difficulties associated with overseeing and enforcing such a regime could, in effect, lead to a myriad of negative unintended consequences. Therefore, it is the CSCRC's view that platform self-regulation and targeted education and intervention, not legislation, may be more effective approaches in the fight against mis- and disinformation.

The rise and rise of the internet has brought a myriad of benefits, enhancing the way we communicate, do business and undertake day-to-day tasks, its rapid unregulated evolution has resulted in significant unintended consequences, including a boom in the spread of mis- and disinformation. As noted by the OECD, the adoption of online social movements has facilitated "the spread of mis- and disinformation, contributed to undermining the role of traditional information gatekeepers … Anyone can be both a producer and a consumer of information, and anybody with an internet connection has the potential to engage with and influence public debates".[2] Hence, there is a clear need to ensure online platforms play a key role in preventing the dissemination of mis- and disinformation, which have the potential to erode our democracy, social systems and economy, and ultimately have a deleterious impact on our way of life.

However, while the need to prevent the spread of mis- and disinformation cannot be understated, a careful line must be drawn between protecting Australians from associated harms, the right to freedom of expression and maintaining the principles of a free and open internet. Frank discourse is a cornerstone of our democracy and, while there must be limits to freedom of expression, which are clearly defined in the UN's International Covenant on Civil and Political Rights (ICCPR),[3] regulating platforms to prevent the dissemination of mis-

[1] Australia on a path to polarisation: Edelman Trust Barometer 2023 | Edelman Australia
[2] 1. Redefining the role of public communication in an evolving information ecosystem | OECD Report on Public Communication : The Global Context and the Way Forward | OECD iLibrary (oecd-ilibrary.org)
[3] Right to freedom of opinion and expression | Attorney-General's Department (ag.gov.au)

and disinformation may, in effect, become 'whack-a-mole' in nature. Ultimately, while such an approach could result in financial penalties for digital platforms, it would not resolve the underlying causes of mis- and disinformation, nor prevent its spread online. Therefore, online education and intervention as opposed to regulation may be a more effective way of targeting the root causes of mis- and disinformation.

In this submission, the Cyber Security Cooperative Research Centre (CSCRC) addresses several key aspects of the proposed the Bill, including:

- definitions of 'misinformation' and 'disinformation';
- the definition of 'serious harm'; and
- challenges of enforcing the proposed regime.

In addition, the CSCRC explores alternatives to regulation, namely:

- Inoculation theory; and
- Platform self-governance via algorithmic monitoring and targeted intervention.

Definitions of 'misinformation' and 'disinformation'

The key distinguishing feature between mis- and disinformation is intent, which is reflected in the definitions used in the Exposure Draft. However, the CSCRC submits the proposed definitions are overly broad in scope and lack clarity, which could lead to difficulty in their interpretation. For example, no examples of what would and would not be classified as mis- and disinformation – beyond excluded content – is provided within the Bill's Guidance Note.[4]

While the rationale for making the definitions broad is understandable given the boundaries between mis-and disinformation are not clear-cut and intent is inherently hard to determine, this approach is also problematic. This difficulty has been highlighted by the European Union's (EU) High-Level Expert Group on Fake News and Disinformation, which observed both mis-and disinformation encompass a spectrum of information types: "This includes relatively low-risk forms such as honest mistakes made by reporters, partisan political discourse, and the use of click bait headlines, to high-risk forms such as for instance foreign states or domestic groups that would try to undermine the political process … through the use of various forms of malicious fabrications, infiltration of

[4] Communications Legislation Amendment (Combatting Misinformation and Disinformation) Bill 2023 — Guidance Note (infrastructure.gov.au), PP 12-13

cybersecuritycrc.org.au

grassroots groups, and automated amplification techniques". [5] Given this spectrum, consideration could be given to developing a tiered system to measure the seriousness of mis- and disinformation to help reduce definitional ambiguity, with the most serious tier being malicious disinformation aimed at undermining Australia's democratic processes and institutions. If such a model were to be considered, the CSCRC believes take down notices, which are not included in the proposed regime, could be an effective solution to the removal of such disinformation materials.

Also problematic in the Australian context is the lack of consistency in defining mis- and disinformation across various parts of government. For example, the Department of Foreign Affairs and Trade (DFAT) definitions are narrower in scope, with misinformation defined as:

- "the creation and dissemination of wholly or partly false information, spread unwittingly, by error or mistake. Such information has the potential to mislead or deceive but is neither created nor transmitted with the intention of doing so or causing harm".[6]

And disinformation defined as:

- "as the intentional creation and dissemination of wholly or partly false and/or manipulated information that is intended to deceive and mislead audiences and/or obscure the truth for the purposes of causing strategic, political, economic, social, or personal harm or financial/commercial gain".[7]

However, the Australian Electoral Commission defines misinformation as "false information that is spread due to ignorance, or by error or mistake, without the intent to deceive", and disinformation as "knowingly false information designed to deliberately mislead and influence public opinion or obscure the truth for malicious or deceptive purposes".[8]

While all similar, these definitional have the potential to impact interpretation, especially in law. Therefore, the CSCRC submits that definitional harmonisation for the terms 'misinformation' and 'disinformation' is essential across the whole of government to reduce confusion and improve clarity.

---

[5] Final report of the High Level Expert Group on Fake News and Online Disinformation | Shaping Europe's digital future (europa.eu)
[6] Disinformation & Misinformation | Australia's International Cyber and Critical Tech Engagement (internationalcybertech.gov.au)
[7] Disinformation & Misinformation | Australia's International Cyber and Critical Tech Engagement (internationalcybertech.gov.au)
[8] eiat-disinformation-factsheet.pdf (aec.gov.au)

Definition of 'serious harm'

The notion of 'serious harm' is difficult to define, as it is subjective and notoriously ambiguous. And in the context of serious harms resulting from mis- and disinformation, it becomes even more complex.

Physical harms caused by mis- and disinformation are more quantifiable than other more abstract harms, such as harm to the Australian environment or economy. For example, there is empirical research to support the hypothesis that exposure to misinformation about COVID-19 has been linked to the ingestion of harmful substances[9] and there are causal links between online misinformation and the propensity to commit real-life hate crimes.[10] However, other harms defined in the in the Exposure Draft, like harm to the integrity of democratic processes or harm to the environment, may also not be immediately evident and could take years to evolve, may take significant time to quantify, may be unverifiable and, if challenged in a court, may be difficult to substantiate.

Under the proposed Bill, the determination of 'serious harm' is dependent on a range of criteria covered in subclause 7(3). The CSCRC submits this range of criteria is overly broad and lacks clarity and, furthermore, would in practice be difficult or impossible to verify. For example, verification of the author of mis- or disinformation (especially from an overseas location) and the purpose of its dissemination may be difficult or impossible to ascertain, with attribution and intent in the digital domain often difficult to conclude. Furthermore, while the table outlining types of harm and how they may reach the threshold of serious harm is helpful, it is not exhaustive. Nor does it account for a range of other less clear-cut scenarios which could arise as a result of the Bill's passage.

Challenges of enforcing the proposed regime

As noted by the World Economic Forum, "regulating industry efforts to stem harmful content is not straightforward because of the difficulty in assigning responsibility and the potential unintended consequences of legal instruments".[11] Furthermore, "because safety is a determination negotiated in public understandings, it cannot be solved by any one

[9] Misinformation: susceptibility, spread, and interventions to immunize the public | Nature Medicine
[10] How hateful rhetoric connects to real-world violence | Brookings
[11] WEF_Advancing_Digital_Safety_A_Framework_to_Align_Global_Action_2021.pdf (weforum.org)

company … the development of safety baselines will need the participation of private corporations, since they know how harm unfolds in technological contexts and how to operationalize solutions".[12] There are other key challenges in relation to enforcement of the proposed regime, notably appropriate and dedicated resourcing of the Australian Communications and Media Authority (ACMA) and extra-territorial application of regulation.

ACMA has a wide remit as a regulator of Australia's communications and media sectors. With the advent of digital communications, this has expanded dramatically. If ACMA were to assume regulatory powers in relation to mis- and disinformation enforcement, this remit would grow further. Hence, if the proposed Bill is enacted, ACMA will need to be properly resourced to handle what the CSCRC expects would be significant additional load. In its most recent annual report, ACMA reported employing 489 staff across four broad key areas of operation.[13] This is not a large workforce and the CSCRC submits that a dedicated division of ACMA would need to be established to oversee the regime, which would also require staff with appropriate expertise in investigations, mis- and disinformation and impact of harms.

The CSCRC notes that eight major digital platforms are signatories to the Australian Code of Practice on Disinformation and Misinformation, with the Digital Industry Group Inc (DIGI) publishing their most recent transparency reports in May 2023.[14] The CSCRC is confident that, with the exception of Tik Tok, these platforms have invested significantly in self-regulating mis- and disinformation on their platforms. However, it is important to note there are many smaller platform operators that will be captured by the proposed regime given the definition of 'connective media service' as contained in 4(3a). Furthermore, many of these platforms operate outside of Australia, in practice making enforcement of fines difficult or impossible, especially for platforms that are able to operate with impunity in particular jurisdictions.

Inoculation Theory

Inoculation theory (the theory), which was developed in the US in 1964, is based in a biological analogy of an organism that has been raised in a sterile, germ-free environment and appears robust and healthy but is in reality vulnerable to infection, because it has not

[12] WEF_Advancing_Digital_Safety_A_Framework_to_Align_Global_Action_2021.pdf (weforum.org)
[13] Australian Communications and Media Authority; Office of the Children's eSafety Commissioner Annual Reports 2015–16 (acma.gov.au), P17
[14] TRANSPARENCY | DIGI

cybersecuritycrc.org.au

had the opportunity to develop defensive antibodies.[15] To 'inoculate' it, small exposures to infectious materials are carried out over time to build up its immunity.

The theory presents a framework for pre-emptive mis- and disinformation interventions, comprising two central elements: warning recipients of the threat of misleading persuasion; and identifying the techniques used to mislead or false information that underpins a false argument to help refute further misinformation.[16] It operates on the assumption that through understanding how misleading techniques are applied in the context of spreading mis- and disinformation, individuals are equipped with the cognitive tools to be aware of and reject further attempts at persuasion.

Inoculation is reliant on two mechanisms - motivational threat, which is a desire to defend oneself from manipulation, and prebunking, in which people are exposed to weakened examples of misinformation.[17] According to research, "the threat component forewarns individuals that they may be exposed to a persuasive attack, and refutational pre-emption either entails directly providing individuals with the counterarguments that refute incoming (mis)information, known as passive inoculation, or it actively involves the participant in the generation of those counterarguments, known as active inoculation". [18]

Research also indicates some people are more susceptible to misinformation. For example, older people have been found to be more susceptible to fake news due to factors including cognitive decline and digital illiteracy.[19] Furthermore cognitive biases established via peer group influences, including online communities, have been found to result in people being more likely to believe information from within their social circle, creating an echo chamber for misinformation to thrive.[20] Such findings could support targeted deployment of inoculation techniques at specific groups as a pre-emptive measure.

In recent years, application of the theory has been shown to increase mis- and disinformation detection and facilitate critical literacy.[21] Digital application of the theory via app-based or web-based games has proven promising, with several studies finding that the process of active inoculation through playing online games significantly reduced the

[15] Inoculation theory - Oxford Reference
[16] The psychological drivers of misinformation belief and its resistance to correction | Nature Reviews Psychology
[17] Misinformation: susceptibility, spread, and interventions to immunize the public | Nature Medicine
[18] Technique-based inoculation against real-world misinformation - PMC (nih.gov)
[19] Misinformation: susceptibility, spread, and interventions to immunize the public | Nature Medicine
[20] Biases Make People Vulnerable to Misinformation Spread by Social Media - Scientific American
[21] The psychological drivers of misinformation belief and its resistance to correction | Nature Reviews Psychology

perceived reliability of news that embedded several common online misinformation strategies[22] and conferred psychological resistance against manipulation techniques commonly used in political misinformation.[23] Furthermore, a study of inoculation videos watched by 30,000 people across several platforms including YouTube, found viewing these videos they improved "people's ability to identify manipulation techniques commonly used in online misinformation, both in a laboratory setting and in a real-world environment where exposure to misinformation is common".[24]

Platform self-governance via algorithmic monitoring and targeted intervention

Valid concerns have been raised about algorithmic biases that arise directly from what people search for and see online, which can create mis- and disinformation echo chambers. As noted by academics, while such technologies were designed to select the most engaging and relevant content for individual users, it has also had the effect of serving to reinforce the cognitive and social biases of users, which may make them more vulnerable to misinformation manipulation.[25]

Detection of mis- and disinformation across digital platforms can – and is - achieved through leveraging algorithms, machine-learning models and human analysis. From a regulatory perspective, given that digital platforms have the ability to exercise control over the spread of information through their networks, it makes sense that self-regulation is supported. For this to be an effective strategy, however, public-private partnerships and co-design are essential. To combat the spread of mis- and disinformation, intervention by internet digital platforms is required, supported by the development of counter-campaigns to neutralise mis- and disinformation by governments, which can be deployed online.[26] Such content could include inoculation messaging to help counter the effectiveness of mis- and disinformation campaigns, as well as intervention materials for people whose algorithmic footprint indicates they frequently access mis- and disinformation materials.

Efforts to counter radicalisation, which in many aspects is comparable to countering mis- and disinformation, have benefitted significantly from content detection and removal. The

[22] Fake news game confers psychological resistance against online misinformation | Humanities and Social Sciences Communications (nature.com)
[2323] Breaking Harmony Square: A game that "inoculates" against political misinformation | HKS Misinformation Review (harvard.edu)
[24] Psychological inoculation improves resilience against misinformation on social media | Science Advances
[25] Biases Make People Vulnerable to Misinformation Spread by Social Media - Scientific American
[26] Algorithms can be useful in detecting fake news, stopping its spread and countering misinformation (theconversation.com)

cybersecuritycrc.org.au

Australian Institute of Criminology (AIC) notes that, while time and labour intensive "content removal schemes have demonstrated some success at removing vast quantities of violent extremist content … For example, Facebook removed 9.4 million pieces of Islamist extremism related content between April and June 2018". Furthermore, the AIC highlights the importance of counternarratives, alternative narratives and strategic communications in countering online extremist content, noting "echo chamber effects and the adoption of polarising views can be decreased by making alternative ideas more widely and readily available, impeding the entrenchment and validation of biased perspectives".[27]

The World Economic Forum (WEF) has also highlighted how sophisticated AI technologies can play a powerful role in countering mis- and disinformation via content analytics. Such tools can, through analysis of content, word patterns, syntax and readability, identify if text is computer or human generated. The WEF notes: "Such algorithms can take any piece of text and check for word vectors, word positioning and connotation to identify traces of hate speech. Moreover, AI algorithms can reverse engineer manipulated images and videos to detect deep fakes and highlight content that needs to be flagged".[28]

Similarly, the United Nations (UN) has highlighted the key role analysing data patterns has to play detecting and countering mis- and disinformation, which can heighten the effectiveness of interventions. According to the UN, "analysing digital trends of what populations are saying at different points in time can help not only better understand social sentiment and engagement with extremist narratives but also provide insights on drivers of violent extremism. Data can also help us reach those most affected by telling us where the gaps might be in areas like education, mental health, employment, attitudes towards women, or social cohesion".[29]

Conclusion

The spread of mis- and disinformation is a serious problem facing the Australian community. And, while steps must be taken to counter their spread and negative impacts on the community, the CSCRC does not believe the proposed Bill will effectively serve this purpose. Hence, it is the CSCRC's position that digital platforms should be supported by government in effective self-regulation. This can be achieved through the establishment of a public-private partnership model that creates inoculation materials and counter-

[27] Understanding and preventing internet-facilitated radicalisation (aic.gov.au)
[28] Is artificial intelligence the antidote to disinformation? | World Economic Forum (weforum.org)
[29] Using online data to tackle violent extremism is a risk worth taking… if we're smart about it. Here's how. | United Nations Development Programme (undp.org)

narratives that can be disseminated via digital platforms, including in targeted ways, to lift the digital literacy of the community more broadly. In relation to serious disinformation, which could be quantified using a tiered model, the government could consider the introduction of take down notices. This would help ensure the most serious disinformation materials, notably those that impact on Australia's democratic systems and institutions, are removed from digital platforms.