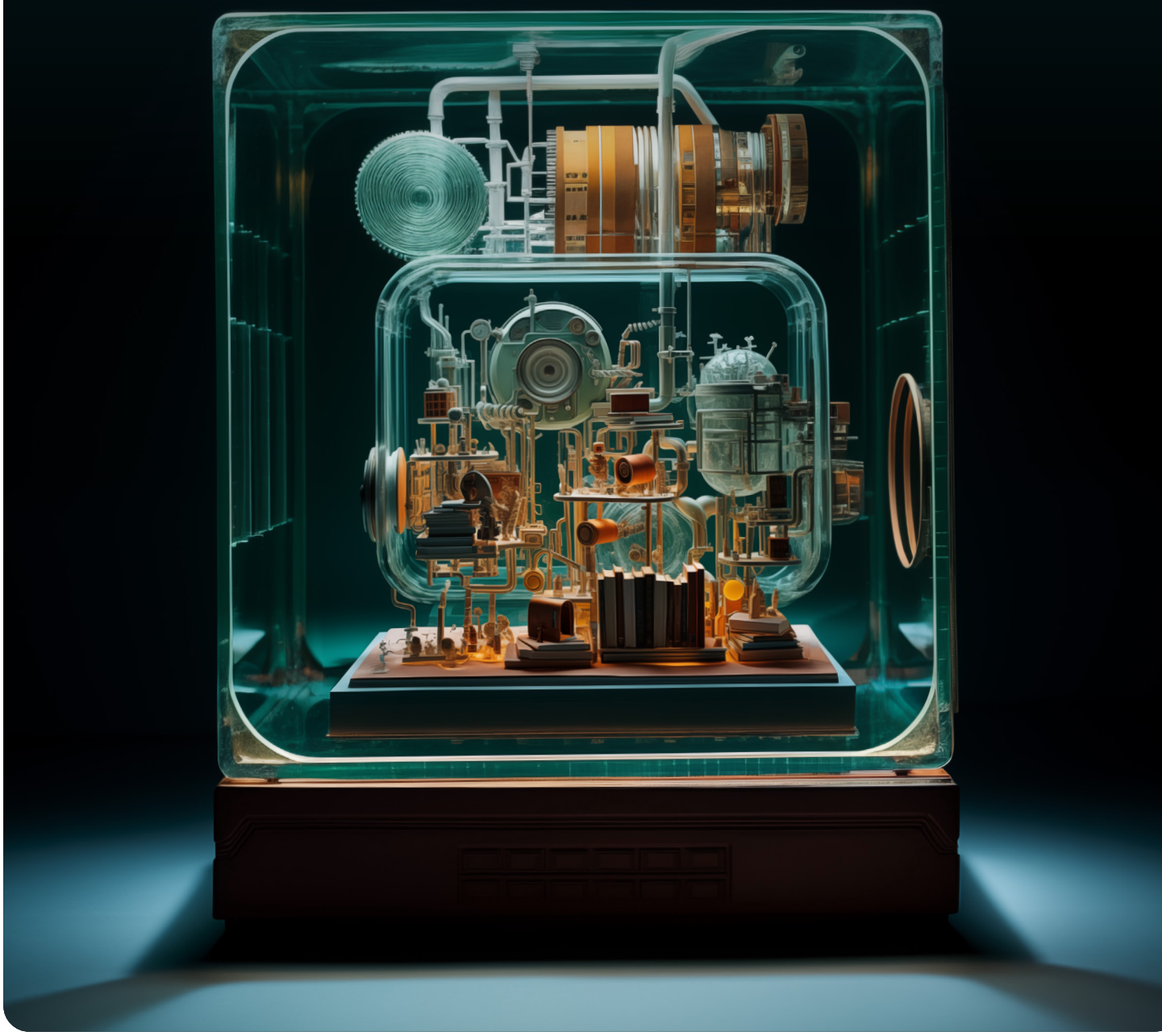
Reset. AUSTRALIA

THE UNIVERSITY OF SYDNEY

Regulating for Transparency: *Transparency Reports in Australia*

Policy briefing | April 2024



Transparency reporting can be a powerful tool but is no silver bullet. Effective transparency reports form one part of a broader transparency framework.

Reset.Tech Australia
2024

Cover image: Artwork created using Midjourney in response to the prompt *"imagine/ a surreal glass book sculpture, futuristic antique style, hasselblad H6D-100c, Sirui 50mm f/1.8 anamorphic 1.33x, --v 5.2*

CONTENTS

Reset.
AUSTRALIA

Reset.Tech Australia is an independent, non-partisan policy research lab committed to driving public policy advocacy, research and civic engagement to strengthen our democracy within the context of technology. We are the Australian affiliate of Reset, a global initiative working to counter digital threats to democracy.

INTRODUCTION

This policy briefing reflects discussions held at a roundtable of 20 experts from academia and civil society in March 2024, where we explored the opportunities and challenges to regulate for transparency in the digital world in Australia. The event was held under the Chatham House Rule, and this briefing presents an overview of the discussion to advance thinking on best-practice transparency reporting and metrics specifically.

The focus on transparency reports is to inform a live policy discussion, with interest to legislators and regulators. The Albanese Government is considering regulatory frameworks for enhancing platform accountability for misinformation and disinformation issues, and the Australian Communications and Media Authority (ACMA) has put out a tender for a consultant to help develop a set of 'metrics' for platform transparency reporting.[1] The tender requires the successful contractor to consult with DIGI, platforms, industry, government and academic experts.

Transparency reporting can be a powerful tool for corporate accountability. But to fully realise this potential, transparency reporting cannot be left as an opt-in exercise of broad, general and unassessable statements. As the European and Australian experience highlights, broad and unhelpful reports emerge under voluntary reporting regimes. Voluntary transparency reporting schemes need legislative and regulatory ballast to ensure that the information provided is sufficiently precise, useful and verifiable. In assessing the limits of voluntary transparency reporting in the lead-up to the *Online Safety Act* (OSA), the UK Government outlined four key issues:

- Those who do report decide what to include in their reports and may not be incentivised to publish certain information that might be useful to users, civil society and government;
- Not all companies that could produce reports choose to do so;
- There is a lack of independent verification of the information provided, which may reduce confidence in the accuracy and value of the data;
- There is significant variation between the reports that different companies currently produce.[2]

Transparency reporting can be a powerful tool but is no silver bullet. Effective transparency reports form one part of a broader transparency framework. The European model has shifted from a reliance on voluntary transparency reports to a more comprehensive regulatory framework where reports sit alongside risk assessments, independent audits, ad repositories and mandated researcher access to platform data. This presents one potential model for Australia to consider.

This memo summarises the discussions held and proposes some recommendations for Australia's policy decision-makers. The event included three provocations, which are summarised below, as well as the broader discussion.

1. TRANSPARENCY REPORTS IN AUSTRALIA

Transparency reports can serve as crucial documents providing insights into platform operations, content moderation, user safety and adherence to regulatory guidelines. However, this is only realised if reporting is done well. In Australia, platform transparency reports released in compliance with the Australian Code of Practice on Disinformation and Misinformation (ACPDM)[3] contain notable gaps that hinder their efficacy.

For example:

Lack of clarity about definitions.

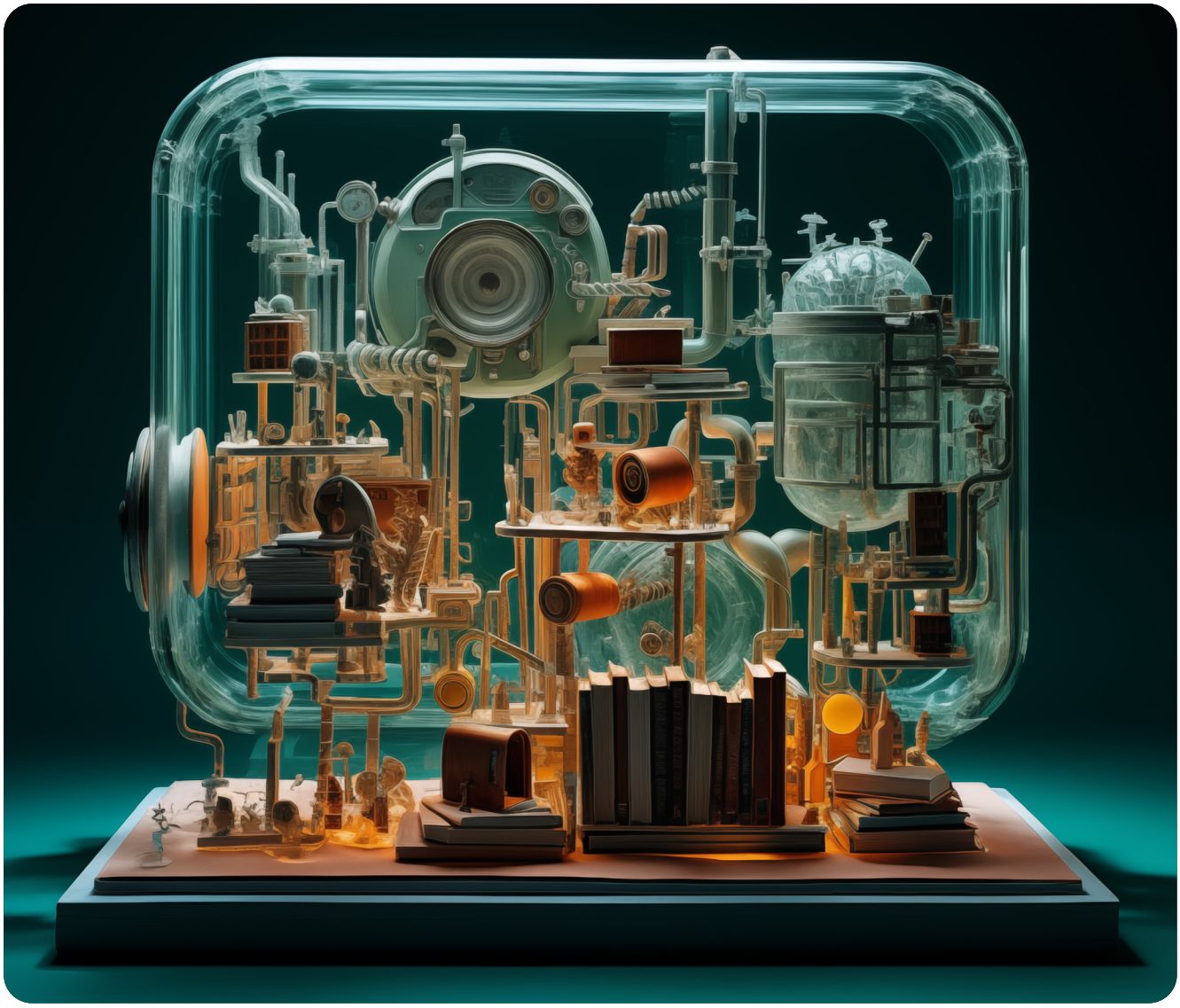
Transparency reports are broadly unclear in at least two key ways. First, when data about violative content is presented in reports, a lack of clarity about the nature of the violation makes the data hard to interpret. For example, when figures about violative content are presented in transparency reports (such as statements like '10,000 posts were removed'), these figures are often 'bundled', and it is unclear which violations they account for (that is, the 10,000 posts could be posts arising from inauthentic behaviour, posts inciting violence, posts threatening electoral integrity or posts proven to be false after fact-checking). Platforms have a range of community guidelines and rules around what content is considered violative, and without specificity, it is hard to understand what the figures in the current transparency reports refer to. Compounding this lack of clarity, variations in guidelines across platforms make comparison difficult if not impossible. What violates guidelines on one platform may be fine on another. Second, reports also talk about 'acting' on content, without providing clear descriptions about what these actions are. Platforms have a range of actions they claim to take against violative content. Platforms can remove it, label it, demote it and so on. Providing bundled figures about actions taken against violative content does not provide sufficient insight.

Lack of clarity about how the content moderation system operates.

Broad and overarching claims are often made about platforms' content moderation systems but with insufficient detail for interrogation. For example, it is unclear what guidelines content moderators are given to enable the evaluation of Australian-specific content. Furthermore, platforms often do not disclose crucial information about their content moderators, and it is unclear how many content moderators are tasked with addressing Australian content. For instance, TikTok's 2022 annual report claimed that the company had 40,000 trust and safety experts,[4] but industry claims suggest their total workforce is only 38,000.[5] It is unclear where these 'outsourced' trust and safety experts – presumably content moderators – work, who they work for and how many of them review Australian-specific content. Additionally, it is unclear whether content moderators are stationed in Australia or abroad, making it difficult to decipher their knowledge of local contexts. Sometimes it is unclear what third-party fact-checkers platforms use.[6] It is often unclear how fact-checkers work, which content they prioritise for checking and how platforms, in turn, moderate their content against fact-checkers' findings.

The figures used in transparency reports lack sufficient context to make them intelligible.

The absence of contextual information in transparency reports raises questions about the significance of reported numbers in three ways. First, the 'denominator issue' makes reported figures impossible to understand. For example, platforms will report the amount of violative content they addressed but without providing estimates of the total number of violative posts on a platform or indeed the total number of posts on a platform, making it very difficult to comprehend the significance of a claim like 'we removed 80K videos that violated community guidelines', as it is unclear if this is a significant proportion of violative content or a drop in the ocean. Second, there is a lack of international comparison. If, for example, 80K videos were removed in Australia, but 500K were removed in New Zealand, we might reach different conclusions about the efficacy of platform actions in Australia. Transparency reports in Australia do not currently require this information, making it impossible to know what proportion of violative content is detected and removed. Third, there is no consistency over time, making it difficult to tell if the situation is getting worse or better on a platform.



2. TRANSPARENCY REPORTING FRAMEWORKS IN AUSTRALIA

The current platform transparency reporting framework in Australia takes place within the ACPDM, designed and administered by DIGI, an industry-affiliated body. DIGI administers a process that takes place roughly as follows:

- Platforms prepare a report on their misinformation and disinformation mitigation efforts for a January-December calendar year;
- These reports are submitted to DIGI in the first quarter of the following year;
- An independent expert reviews the reports;
- Reports are released publicly in May.

Templates for the reports are set out in Appendix 2 of the ACPDM.[7]

There are two key issues with the current reporting framework under the ACPDM:

The transparency reports that emerge from this process are often low quality.

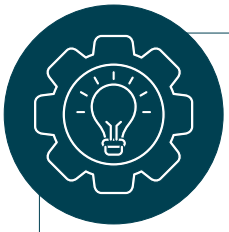
As discussed above,[8] reports that emerge from the current framework lack clarity and specificity and often read as public relations documents rather than comprehensive, accountable statements on trust and safety efforts.

The framework for producing reports under the ACPDM does not drive quality reporting. Platforms have been known to justify their lack of clarity by citing Best Practice Guidelines that encourage brevity over comprehensiveness. Furthermore, platforms have implied that under these Best Practice Guidelines, the objective of their transparency reports is to meet the minimal narrative requirements as laid out by DIGI rather than to provide coherent insights to the public or public oversight bodies.

However, low-quality transparency reporting is not an issue unique to digital regulation, and there are potentially a lot of takeaways from the environmental movement. In this space, companies have moved from voluntary, 'corporate social responsibility' driven reporting to mandatory and standardised reporting (see Figure 1).

The content of the reports is difficult to challenge under the current framework.

Formal engagement avenues with signatories via the ACPDM are limited. External engagement with signatories and DIGI is generally through the Complaints mechanism. The lack of specificity in the transparency reports themselves means that of the limited complaints avenues available, they are more likely confined to interpretative disagreements over *how activities are represented* rather than *the efficacy of those activities themselves*. For example, if a statement in a platform's transparency report is found to be materially false, it constitutes a material breach.[9] Yet it is almost impossible to reach a factual finding of *falseness* without a) the use of clear and accessible metrics, b) the availability of data benchmarks and c) appropriate data access to test platform claims. It is also unclear how this applies to claims that are misleading or unclear, creating a paradox where complaints cannot be made about transparency reports for not being transparent.



A more comprehensive, public-facing and accountable framework is required.

Transparency reports will not be effective on their own. While there is an urgent need for better transparency reports in Australia, they should not be the single mechanism to provide public transparency. We have identified five public transparency measures that could be required in a revised Australian framework:

1. Annual transparency reports. To improve quality, transparency reports should be authored by platforms but must address a clear set of metrics provided by regulators. Additionally, an ombudsman mechanism should be established to allow the public and civil society to raise complaints about the contents of these reports.

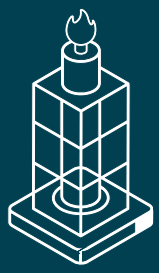
2. Annual risk assessments. Authored by platforms, these documents capture the identified systemic risks and mitigation measures undertaken by a platform. These risk assessments serve many important purposes which are not addressed in this paper, but also have a public transparency function. While they may initially be submitted to regulators, a summarised version should be released to the public after a period of time.

3. Annual independent audits that scrutinise both risk assessments and transparency reports.[10] Because platforms hold the most knowledge about the risks occurring on their platform, the mitigation measures they employ and their impacts, they are best placed to author transparency reports and risk assessments. However, it is not an effective strategy to allow platforms to 'mark their own homework'. To avoid this, platforms must be required to undertake independent audits with evidentiary testing of claims made in their transparency reports and risk assessments. We note that currently the reports produced under the ACPDM are reviewed by an independent expert, but this process does not appear to include significant data testing or reach the threshold of an audit. These documents should become publicly available.

4. Ad repositories that make paid-for advertising visible and searchable, including information about the advertisers. These repositories create searchable archives of advertising and, for political content, provide details about targeting categories and expenditures.

5. Access to public interest data for vetted researchers and research organisations. Requirements for researcher access presume that platforms will provide non-commercial researchers with necessary data for public interest research. For example, in the DSA, this is operationalised as a requirement to provide access to data to 'vetted researchers' upon their request for the purpose of researching systemic risks arising on the platform.[11]

These five public transparency and accountability mechanisms need to exist alongside strong investigative powers, allowing regulators to request information and, critically, act upon it in meaningful ways.



Lessons learnt from the environmental sector:

Voluntary reporting on greenhouse gas and carbon emissions began in the late '90s, largely driven by corporate social responsibility objectives, with individual companies choosing to report their carbon emissions in non-standardised ways. As the urgency of the issue became increasingly apparent, jurisdictions such as the EU intervened by enacting regulations mandating large companies to report carbon emissions.[12] Similarly, in 2023, the Australian Government announced its commitment to introduce mandatory reporting, with clear metrics and targets:

'... the Government will introduce standardised, internationally-aligned reporting requirements for businesses to make disclosures regarding governance, strategy, risk management, targets and metrics – including greenhouse gasses'.[13]

Figure 1: The move from voluntary emissions reporting to mandatory, standardised reporting

Photo credit: Tim van der Kuip on Unsplash

3. THE EU APPROACH TO PUBLIC TRANSPARENCY

An effective transparency framework creates meaningful tools for public oversight, but this requires careful construction. The experience in the European Union is potentially a useful case study, having experienced the 'transparency theatre' of unclear transparency reports under a voluntary code and now moving towards effective frameworks under the *Digital Services Act* (DSA).[14]

The DSA is a relatively new piece of legislation and has undergone staged implementation, being in full force only since February 2024. Although it is too early to assess its full impact, the DSA includes five significant transparency features: annual transparency reports, risk assessments, independent audits, ad repositories and researcher access. Some of these features are beginning to make an impact.

For example:

- **The first round of transparency reports under the DSA has been released.**[15] These detailed reports include data on content takedowns and insights into the operations and resources allocated to content moderation systems. For instance, they highlight the human resources invested in content moderation systems, revealing significant issues such as the lack of resources in smaller markets like Eastern Europe.[16] Given the region's susceptibility to Kremlin-backed disinformation networks,[17] both civil society and regulators have taken a keen interest in these findings. This data is also crucial for Australia, which has over 170 spoken languages.
- **The first platform risk assessments have been submitted to the European Commission.** These assessments identify platform risks, including those related to broader harms as defined in the DSA (see Figure 2). They also outline the activities and mitigation measures they are putting in place around these. Reporting on each of these broader harms covers the risks that misinformation and disinformation pose. These are not yet public, regulators are actively reviewing them to inform enforcement strategies. Summaries are expected to be released to the public over time.
- **The 'researcher access' scheme is emerging,** providing an additional layer of scrutiny. Although still in its early stages, major platforms have released applications and details about the process. While it remains unclear how much friction may be involved in these processes, the initial step of announcing the schemes and applications has been completed.[18]

Although transparency reports are the prime rib of the DSA, past experiences with the EU's voluntary Code of Practice on Disinformation[19] highlight the importance of not solely relying on them. This code also mandated transparency reports, but like Australia's, they were often unintelligible. They were excessively long and lacked information necessary for insights into the scale of the problem or the efficacy of platform actions. The DSA's more comprehensive transparency framework builds on a wealth of experience, from which the Australian context can hopefully also learn.

Article 34 of the DSA mandates platforms to mitigate risks related to:

- Illegal content;
- Negative impacts on the exercise of fundamental rights such as dignity, privacy and political freedoms, as outlined in the European Charter;
- Civic discourse and electoral processes, as well as public security;
- Gender-based violence, public health, children's wellbeing and significant negative consequences to individuals' physical and mental wellbeing.

The role of misinformation and disinformation in exacerbating these risks will be addressed within this framework.

Figure 2: Details about the risks platforms must mitigate under Article 34 of the DSA.[20]

Photo credit: Maximalfocus on Unsplash

DISCUSSION

The discussion focussed on three overlapping areas.

Disinformation as a digital risk

Misinformation and disinformation are not isolated risks in the digital world, and understanding their interconnectedness through two lenses can be helpful in envisioning appropriate regulatory responses.

First, they are connected to and often an integral aspect of other digital risks, such as algorithmic distortion, where content recommender systems prioritise engaging content over truth,[21] leading to the rapid spread of misinformation and the formation of filter bubbles.[22] Additionally, there is a risk inherent in gamified engagement features, where users are encouraged to maximise likes, comments, etc.[23] These issues arise due to the problematic business model of many digital platforms, which prioritise advertising revenue and user engagement at the expense of social values.

The discussion noted that Australia is unique in treating misinformation and disinformation as a standalone issue disconnected from other digital risks. For example, the DSA does not position disinformation as a standalone risk; rather, it is considered a contributing factor to fundamental risks outlined in the DSA, such as risks to fundamental rights, electoral integrity and minors. Additionally, the DSA requires platforms to mitigate risks to public health, requiring a focus on how disinformation exacerbates this risk and the steps taken to mitigate it. Addressing disinformation as a standalone risk is politically contested because it focuses solely on content rather than the broader risks associated with platform content handling.

Second, misinformation and disinformation serve as instigators of interconnected online harms, ranging from mental health issues among young people[24] to online harassment.[25] While these harms are often addressed under the lens of online safety in Australia, enhancing transparency mechanisms around misinformation and disinformation can enable a focus on addressing the root causes of these harms rather than just the effects. Pursuing transparency for its own sake is not productive or helpful; instead, transparency reports should be viewed as tools to comprehend the emergence of online harms and to shape effective remedies.

The OSA and its connections to this issue should be considered, as the review of the OSA presents a prime opportunity to further enhance corporate accountability. If the OSA is intended to be Australia's primary regulation aimed at establishing an upstream, systemic architecture to manage digital risks, its role in fostering conditions for transparency and accountability must not be overlooked.[26]

There were discussions about how Australia could be more ambitious than merely advocating for 'transparency' alone and instead create a regulatory regime that addresses the most egregious aspects of platforms' business models. This entails focusing on a broader array of digital risks beyond disinformation alone. The next step might involve considering what regulations are necessary to incentivise a healthier platform business model and what characteristics of online information ecosystems would bolster democracy. This could include aspects such as media plurality, free speech and non-discrimination.

Learning from and going beyond the *Digital Services Act*

The European experience and emerging evidence regarding effective transparency measures could guide Australia to surpass the DSA. For instance, early evidence from transparency reports in Europe suggests that how platforms report 'time' can be problematic. Some platforms report in hours, while others use partial days, making it challenging to comprehend and compare reports across different platforms. Australia could refine these metrics and iterate on them to establish a more robust transparency regime.

A combination of information sources could assist Australia in developing a world-class transparency regime. This would entail data regarding platforms' content and takedown rates, along with descriptions of risks, mitigation measures and associated metrics. Although Australia may not currently have proposals for mandatory risk assessments regarding disinformation, transparency measures could still promote visibility into mitigation measures and business practices.

The discussion highlighted similarities between Australia's current situation and the EU's experience around 2018. In Australia, our current voluntary code appears ineffective at fostering meaningful transparency, sparking discussions about regulatory intervention. This echoes the EU's experience, where the Code of Practice on Disinformation[27] was initially developed in 2018 but fell short of achieving the desired changes advocates had hoped for. However, a key difference could lie in the broader visibility of the issue. The size of the EU and the presence of global clusters of excellence within its member states created a strong internal demand for a comprehensive and effective transparency regime.

The role of civil society in shaping the transparency framework

In the EU, the transparency framework was significantly influenced by civil society (spanning non-government and not-for-profit organisations, charities, research organisations, and academics). For instance, the academic community ensured that the researcher access provisions were not diluted, and civil society established well-organised networks to counterbalance the deep-pocketed lobbying power of platforms. There was a discussion about how Australian civil society could be empowered to fulfill a similar role within this evolving debate.

There was a widespread belief that the current transparency reports were not beneficial to civil society in assessing or understanding the harms occurring online. Simultaneously, there was recognition that broader transparency provisions, such as researcher access, could be valuable.

RECOMMENDATIONS

Three key recommendations emerged from this discussion.

1. **Transparency must be considered as an integral pillar of effective digital regulation.**

   As Australia progresses towards comprehensively addressing digital risks, several regulatory reforms are underway. These include proposals for a Combatting Misinformation and Disinformation Bill,[28] a review of the *Online Safety Act*,[29] the implementation of the *Privacy Act* review and the ongoing Digital Platforms Services Inquiry 2019–2025.[30] The significance of effective transparency frameworks in all of these policy domains should be acknowledged.

2. **Regulatory mandates for publishing transparency reports must include a detailed set of metrics.**

   If requirements to produce transparency reports are included in the Combatting Misinformation and Disinformation Bill or as part of the *Online Safety Act* review, specific metrics must be stipulated. Without clear requirements, annual reports will remain a part of 'transparency theatre'. A clear set of metrics that provide clarity about definitions, the operations and the effectiveness of content moderation systems, along with enough context to make figures intelligible, must be provided to platforms. Appendix 1 provides some indicative examples regarding misinformation and disinformation. Civil society needs to be engaged in the discussion about finalising these metrics.

3. **Broader transparency frameworks should be considered for the Combatting Misinformation and Disinformation Bill and as part of the *Online Safety Act* review.**

   Detailed, contextualised transparency reports need to be seen as one part of a broader transparency framework. While they can be a very effective tool for increasing transparency, they will not deliver meaningful transparency alone. Additional measures, such as regulator-facing risk assessments that eventually become public, independent audits, ad repositories and researcher access, are required. These 'public' transparency measures must accompany robust investigative powers for regulators.

ACKNOWLEDGEMENTS

This briefing paper reflects the expertise of those who contributed to the roundtable and paper. Attendance does not necessarily imply endorsement.

This includes:

› Aruna Anderson, Reset.Tech Australia
› Alice Dawkins, Reset.Tech Australia
› Alice Drury, Human Rights Law Centre
› David Mejia-Canales, Human Rights Law Centre
› Dr Rys Farthing, Reset.Tech Australia
› Felix Kartte, Independent consultant
› Rita Jabri Markwell, AMAN
› Professor Uri Gal, University of Sydney

All errors and omissions rest with Reset.Tech Australia.

This report was prepared with support from the Susan McKinnon Foundation.

Susan McKinnon Foundation

APPENDIX 1:

Indicative metrics for transparency reports relating to misinformation and disinformation

Some potential metrics developed by Reset.Tech Australia for a prescriptive transparency report are presented below as a 'starting guide' and are neither comprehensive nor exhaustive.

Subject area	Example metrics
Volume and response to regulator orders and other legal requirements	Number of 'take down' orders issued by regulators (and other agencies where relevant); median, average and max time in days to respond to these; and final response
	Number of notices received regarding IP, defamation, privacy and illegal content notifications received from Australian end-users; median, average, and max time to respond to these; and final response
	Notices processed using automated means
	Data about the number of out-of-court settlements made
Content moderation metrics, including assessment of impact on Australian businesses and pages	Number of organic content measures (i.e. how much content they proactively detected) that violated their community guidelines; by violation type (e.g. violated self-harm policy or was fact-checked misinformation); amount detected by automated means; amount detected by human moderators; median, average and max time in days to detect these; and final response. Examples of content could be helpful to add to the narrative
	Business specific metrics, e.g. · Number of organic business entity measures (i.e. how many Australian business accounts were removed and restricted as a result of organic content moderation) · Number of organic entity measures (i.e. how many Australian pages or products were removed and restricted as a result of organic content moderation)
	Number of 'trusted-flagger' content measures (i.e. how much content was acted on by a platform as a result of Australian fact-checkers or trusted flaggers); amount reported to platform; by violation type; amount subsequently detected by automated means; median, average and max time in days to detect these; response; number of challenges against response; final outcome
	Indicators of accuracy and error rates for automated review processes, both for organic detection and following user reporting
	Human resources dedicated to content moderation, including information about: number located within Australia; number dedicated to Australian content or addressing reports from Australian end-users; qualifications and training; support; volume of work (i.e. how much content per hour they are required to review); and languages addressed
Measures against misuse	Number of accounts suspended or deleted and why; number of challenges and final outcome
Local usage data	Number of Australian end-users monthly, including breakdowns by under 18 and over 18; median, average and max time spent on platforms in hours per day, broken down by under 18 and over 18

ENDNOTES

1 Aus Tender 2024 *Tender to establish Contract for Analysis of metrics to assess digital platform effectiveness under voluntary, self-regulatory* https://www.tenders.gov.au/Atm/ShowClosed/32c0f214-5a50-4cf9-98e1-b2f72aa54ddc?PreviewMode=False

2 HM Government 2020 *The Government Report on Transparency Reporting in relation to Online Harms* https://assets.publishing.service.gov.uk/media/5fd8b0d9e90e071be3500aa8/The_Government_Report_on_Transparency_Reporting_in_relation_to_Online_Harms.pdf, 5.

3 Digi 2022 *Australian Code of Practice on Disinformation and Misinformation* https://digi.org.au/disinformation-code/

4 TikTok 2022 *Annual Transparency Report* https://digi.org.au/wp-content/uploads/2023/05/TikTok-2022-Annual-Transparency-Report.pdf

5 Ch Daniel 2024 *'TikTok Users and Growth Statistics (2024)' SignHouse* https://www.usesignhouse.com/blog/tiktok-stats

6 Note, some information about this is publicly available, for example:

- Meta 2024 *Where We Have Fact Checking* https://www.facebook.com/formedia/mjp/programs/third-party-fact-checking/partner-map
- Arjun Narayan Bettadapur 2020 *'TikTok partners with fact-checking experts to combat misinformation' TikTok* https://newsroom.tiktok.com/en-au/tiktok-partners-with-fact-checking-experts-to-combat-misinformation

A full list of Australian operating fact-checkers can be found at IFCN 2024 *Verified signatories of the IFCN code of principles* https://ifcncodeofprinciples.poynter.org/signatories

7 See Appendix 2, DIGI 2022 *Australian Code of Practice on Misinformation and Disinformation* https://digi.org.au/wp-content/uploads/2022/12/Australian-Code-of-Practice-on-Disinformation-and-Misinformation-FINAL-_-December-22-2022.docx.pdf

8 See also Reset.Tech Australia 2023 *Policy briefing: Misinformation and disinformation regulatory frameworks* https://au.reset.tech/news/policy-briefing-misinformation-and-disinformation-regulatory-frameworks/

9 DIGI 2021 *Terms of reference for Complaints Facility and Complaints Sub-committee: The Australian Code of Practice on Disinformation and Misinformation* https://digi.org.au/wp-content/uploads/2021/10/DIGI-TOR-for-Complaints-Facility-and-Complaints-Sub-committee-_-ACPDM-_-FINAL-NE-1.pdf (Glossary section j) part (iii)).

10 This could build on the European model for independent oversight. See: European Commission 2023 *Delegated Regulation supplementing Regulation (EU) 2022/2065 of the European Parliament and of the Council, by laying down rules on the performance of audits for very large online platforms and very large online search engines* https://digital-strategy.ec.europa.eu/en/library/delegated-regulation-independent-audits-under-digital-services-act

11 EU 2022 *Digital Services Act* https://eur-lex.europa.eu/eli/reg/2022/2065/oj, Article 40

12 European Parliament 2022 *Sustainable economy: Parliament adopts new reporting rules for multinationals* https://www.europarl.europa.eu/news/en/press-room/20221107IPR49611/sustainable-economy-parliament-adopts-new-reporting-rules-for-multinationals

13 Department of the Treasury 2023 *Climate-related financial disclosure Consultation paper Australian Federal Government* https://treasury.gov.au/sites/default/files/2023-06/c2023-402245.pdf

14 EU 2022 *Digital Services Act* https://eur-lex.europa.eu/eli/reg/2022/2065/oj

15 For a tracker to compare these, see Gabby Miller 2023 *Tracking the First Digital Services Act Transparency Reports Tech Policy Press* https://www.techpolicy.press/tracking-the-first-digital-services-act-transparency-reports/

16 Global Witness 2023 *How Big Tech platforms are neglecting their non-English language users* https://www.globalwitness.org/en/campaigns/digital-threats/how-big-tech-platforms-are-neglecting-their-non-english-language-users/

17 David Klepper 2022 *Russia's information war expands through Eastern Europe The Associated Press* https://apnews.com/article/russia-ukraine-nato-bulgaria-misinformation-eastern-europe-0e87db7fef9263a465d6cf40d3287efe

18 For those who are interested, current researcher access schemes can be found at:

- Meta n.d. *Meta Content Library and Content Library API* https://somar.infoready4.com/#freeformCompetitionDetail/1910793
- X n.d. *X DSA Researcher Application* https://docs.google.com/forms/d/e/1FAIpQLSdo0O-D6Kxa3cV4g1JLz2T_0Sk3hdEnTdv8dJmibagCnzJ7kg/viewform
- AliExpress n.d. *AliExpress Open Research & Transparency Application for access to publicly accessible information by researchers* https://yida.alibaba-inc.com/o/research/api?spm=a2g0o.
- TikTok n.d. *Research API* https://developers.tiktok.com/products/research-api/
- Google n.d. *Google Researcher Program Application* https://requestrecords.google.com/researcher/form
- LinkedIn 2023 *Researcher Access* https://www.linkedin.com/help/linkedin/answer/a1645616?src=or-search&veh=www.google.com%7Cor-search
- Snapchat n.d. *Researcher Data Access Instructions* https://values.snap.com/en-GB/privacy/transparency/researcher-access
- Booking.com n.d. *Booking.com Researcher Data Use Policy* https://dsarequests.contentintegrity.booking.com/hc/en-gb/articles/23861502709396-Booking-com-Researcher-Data-Use-Policy

19 European Commission 2022 *EU Code of Practice on Disinformation* https://digital-strategy.ec.europa.eu/en/library/2022-strengthened-code-practice-disinformation

20 Article 34 EU 2022 *Digital Services Act* https://eur-lex.europa.eu/eli/reg/2022/2065/oj

21 Peter Dizikes 2018 *Study: On Twitter, false news travels faster than true stories Massachusetts Institute of Technology* https://news.mit.edu/2018/study-twitter-false-news-travels-faster-true-stories-0308

22 See, for example, Reset.Tech Australia 2024 *Youth Radicalisation on YouTube Shorts: A live experiment* https://au.reset.tech/news/youth-radicalisation-on-youtube-shorts-a-live-experiment/ and Reset.Tech Australia 2022 *Algorithms as a weapon against women: How YouTube lures boys and young men into the 'Manosphere'* https://au.reset.tech/news/algorithms-as-a-weapon-against-women-how-youtube-lures-boys-and-young-men-into-the-manosphere/

23 These methods are effective and can create risks (see, for example, Mohammad Hajarian, Azam Bastanfard, Javad Mohammadzadeh and Madjid Khalilian 2019 *A personalized gamification method for increasing user engagement in social networks* https://link.springer.com/article/10.1007/s13278-019-0589-3 and 5Rights Foundation 2021 *Pathways: How digital design puts children at risk* https://5rightsfoundation.com/uploads/Pathways-how-digital-design-puts-children-at-risk.pdf)

24 Indeed, a December 2023 poll of 1,008 Australian teenagers aged 15–17 revealed that they identified misinformation as one of the top online risks they faced, tied with online scams as the joint first issue. This perception surpassed concerns about online abuse and exposure to distressing material. For more details, see Reset.Tech Australia 2024 *Response to the Amending Online Safety (Basic Online Safety Expectations) Determination 2022 consultation* https://au.reset.tech/uploads/Basic-Online-Safety-Expectations-Reset.Tech-Submission-Feb-24-pdf

25 See, for example, Brandy Zadrozny 2023 *'A fake tweet spurred an anti-vaccine harassment campaign against a doctor' NBC News* https://www.nbcnews.com/tech/misinformation/fake-tweet-spurred-anti-vaccine-harassment-campaign-doctor-rcna64448

26 There are also rich discussions among civil society around the need for transparency from government activities and requests that may affect the digital environment. See for example, Aman 2023 *Government Transparency* https://www.aman.net.au/?page_id=2132

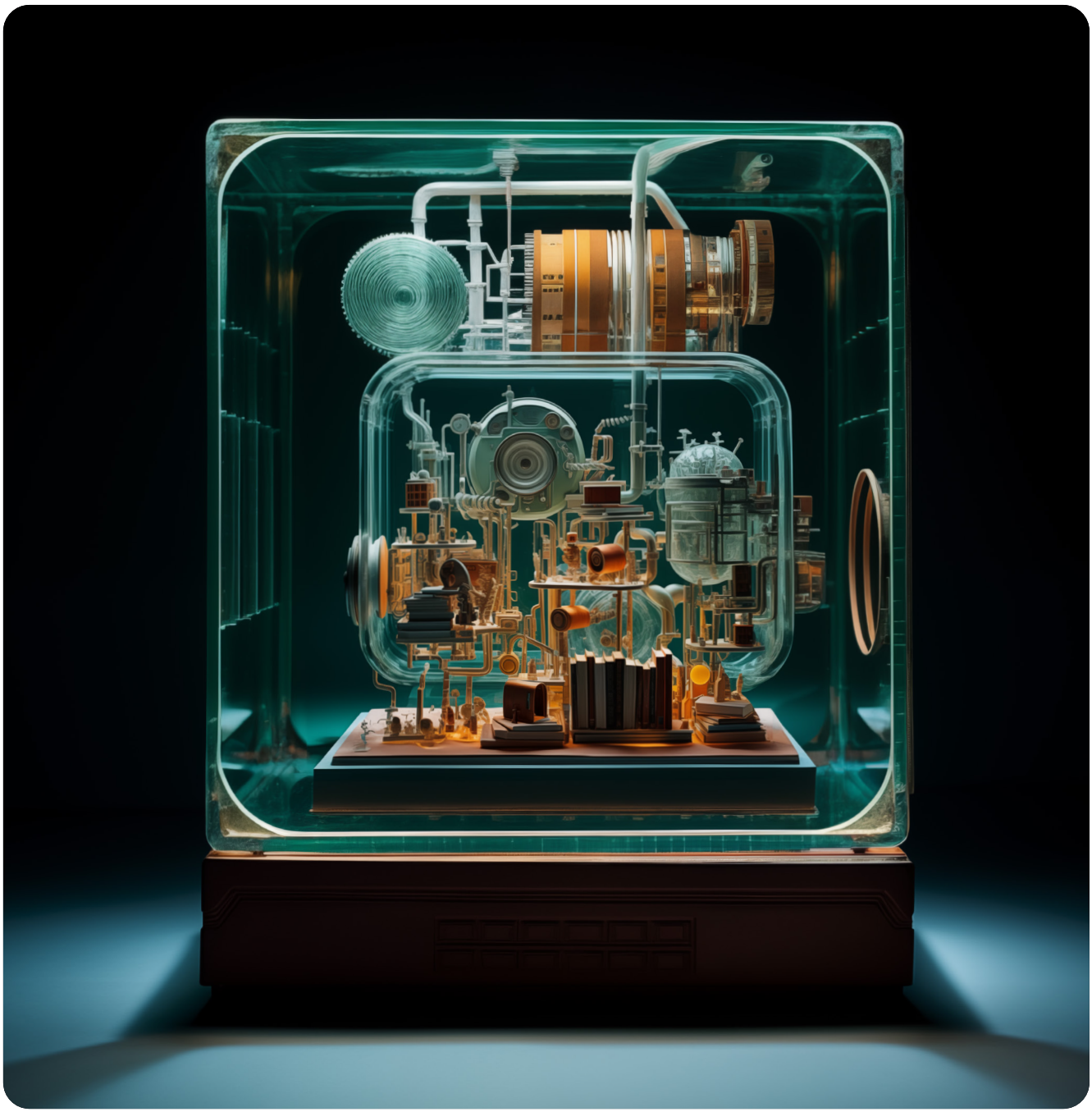
27 European Commission 2022 *EU Code of Practice on Disinformation* https://digital-strategy.ec.europa.eu/en/library/2022-strengthened-code-practice-disinformation

28 Department of Infrastructure, Transport, Regional Development, Communications and the Arts 2023 *Australian Federal Government New ACMA powers to combat misinformation and disinformation* https://www.infrastructure.gov.au/have-your-say/new-acma-powers-combat-misinformation-and-disinformation

29 Department of Infrastructure, Transport, Regional Development, Communications and the Arts 2024 *Terms of Reference – Statutory Review of the Online Safety Act* 2021 *Australian Federal Government* https://www.infrastructure.gov.au/sites/default/files/documents/tor-statutory-review-online-safety-act-2021-8Feb.pdf

30 Australian Competition and Consumer Commission 2020 *Digital platform services inquiry 2020–25* https://www.accc.gov.au/inquiries-and-consultations/digital-platform-services-inquiry-2020-25

Reset.
AUSTRALIA



Artwork created using Midjourney in response to the prompt *"imagine/ a surreal glass book sculpture, futuristic antique style, hasselblad H6D-100c, Sirui 50mm f/1.8 anamorphic 1.33x, --v 5.2*