Reset.
AUSTRALIA

Accountability, the *Online Safety Act* and the *Basic Online Safety Expectations*: **Can safety standards be enforceable?**

Policy briefing | February 2024

A number of amendments are required to the proposals around the *Basic Online Safety Expectations* and the *Online Safety Act* to ensure that Australian end-users enjoy meaningful safety standards with equivalence to our international counterparts.

Reset.Tech Australia
2024

Version 2

CONTENTS

Cover image: *'safety in the clouds'* Artwork created using Midjourney in response to the prompt *"imagine/ a fluffy cloud art installation at sunrise, wide aerial view, people sitting, Hasselblad H6D-100c, Sirui 50mm f/1.8 anamorphic 1.33x, --v 5.2*

INTRODUCTION

This policy briefing reflects discussions from a roundtable of 22 policy experts on January 25th 2024. The event was held under the Chatham House Rule.

The roundtable was prompted by proposals put forward in the *Online Safety (Basic Online Safety Expectations) Amendment Determination 2023*[1] and the proposals to grow and expand Australia's online safety standards, to include:

- New basic expectations regarding user safety in the use of generative AI, recommender systems and user controls;
- Introducing requirements around considering children's best interests in the design and deployment of products, and restricting children's access to class 2 (pornographic) materials online;
- New expectations around the safety impacts of business and resourcing decisions;
- Including basic expectations around content containing online hate speech;
- Improving measures for transparency, such as requiring regular transparency reports, and;
- Introducing requirements for platforms to enforce their terms of use.

While these gradual expansions of expectations may be welcome, there was no clear path noted in the review towards making these standards enforceable. Voluntary safety standards frequently fail to be implemented[2] and there is concern across civil society that these welcome expansions of the Basic Online Safety Expectations (BOSE) may fall short of driving up safety standards in practice.

Alongside this, a broader review of the *Online Safety Act* 2021 has been announced for early 2024, and there is the potential to review the approach of the BOSE, including issues around enforceability.

Against this backdrop, Reset.Tech Australia convened a roundtable to explore the proposals in the BOSE and potential reforms of the *Online Safety Act* from a range of safety and regulatory perspectives and to ask the challenging question: can our safety standards be enforceable?

Three models for regulator-enforced safety legislation provoked the discussion, from the UK regarding the *Online Safety Act* 2023*,* from the EU regarding the *Digital Services Act* 2022 and from Canada regarding their emerging online safety Bill. Discussion of these models and some comparative analysis are summarised in this report.

1. THE *UK'S* ENFORCEABLE SAFETY STANDARDS

The UK's *Online Safety Act* ('UK OSA') was passed in November 2023. It includes duties of care for platforms which fall into three categories:

1. **Content focussed** (illegal content, content that is harmful to children);
2. **Systems focussed** (safety by design/child safety, design principles for children, privacy); and
3. **Other issues, including transparency, redress and political freedoms.**

The UK OSA places the same level of obligations on platforms regarding the content duties, the design duties and transparency and freedom duties. In other words, expectations for content, systems design, and transparency are relatively evenly weighted.

The UK OSA is unique in its even coverage of systems *and* content. While there is an emerging global trend towards focussing on systems, it is important to recognise that some content online is absolutely illegal and inherently harmful (for example, CSAM). Australia's *Online Safety Act*, by comparison, focuses largely on content. The UK has developed an Act that takes a hybrid approach.

The hybrid approach has been generally well-received, and the recognition of illegal and harmful content is popular with the public. Broad and consistent public support for better online protections furnished lawmakers with the ability to pass the legislation even in conditions of political turbulence.

In terms of ensuring compliance, Ofcom—the UK's media and broadcasting regulator—was given additional powers and resourcing to meaningfully enforce the OSA. Ofcom has rapidly scaled up a team of some 300 people dedicated to online safety.

The UK OSA gives Ofcom a range of powers including:

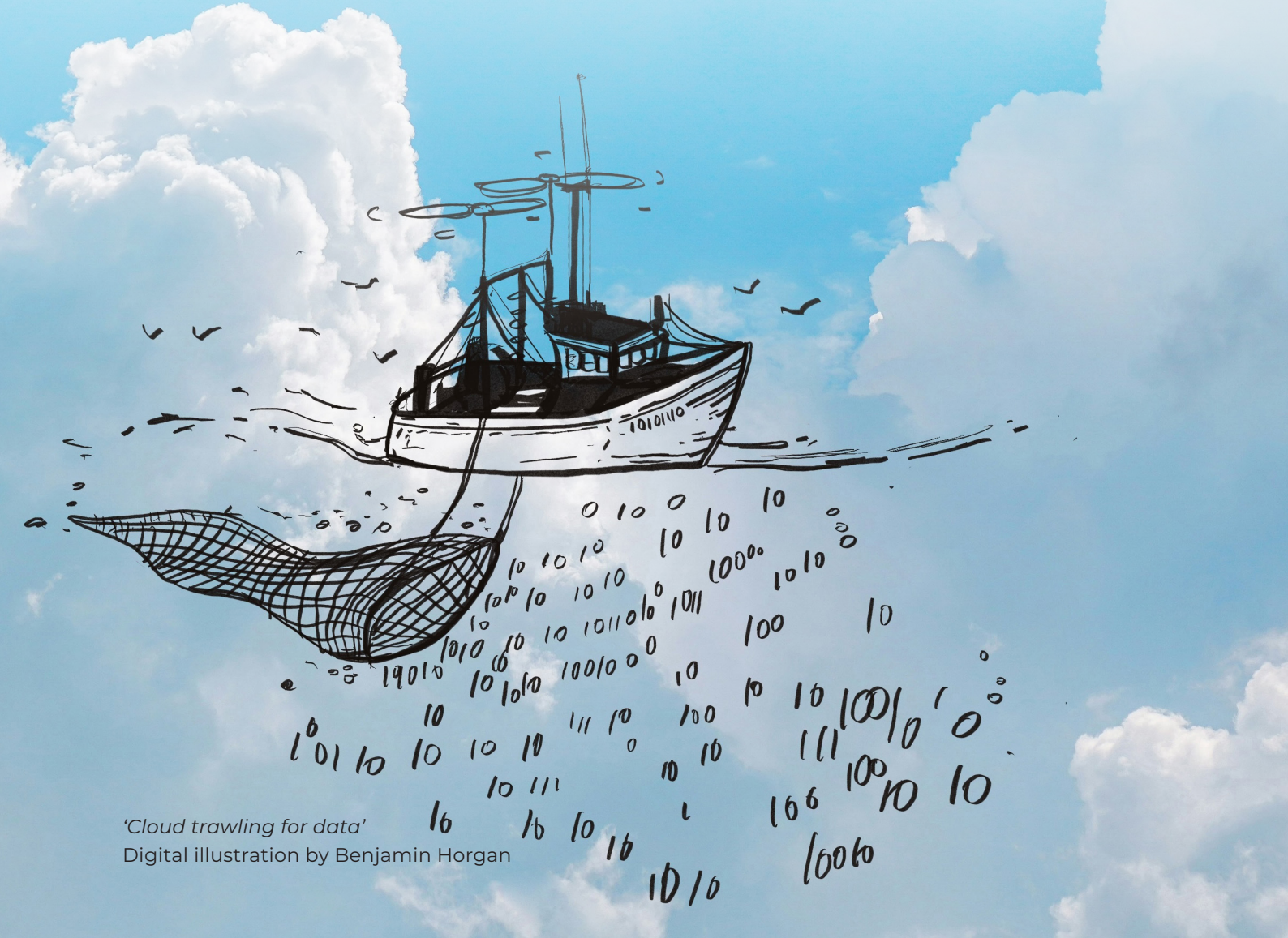
› **The ability to write Codes and guidance.** Part of Ofcom's role is to develop Codes about how the UK OSA should be implemented and understood by industry. This differs from the Australian process where industry writes their own Codes, and these Codes are then considered by the regulator. In this sense, the UK approach involves the regulator drafting would be called 'industry standards' in Australia. Throughout the Code drafting process, Ofcom has obligations to consult a wide variety of stakeholders, but ultimately holds the pen on what the Codes look like in the final instance. Ofcom also has the ability to write guidance on various issues, such as violence against women and girls. This guidance is not as legally binding and does not have the same teeth as the Codes do, but Ofcom are empowered to draft it where they feel it is necessary.

› **Ensuring risk assessments are undertaken and risks are appropriately mitigated.** Platforms are expected to conduct risk assessments against the Codes drafted by Ofcom, and to introduce risk mitigation measures. Ofcom can review these risk assessments and compel redress where they believe mitigation measures are not adequate. By way of comparison, under Australia's OSA, the Office of the eSafety Commissioner has no powers to compel redress, i.e. the regulator cannot demand that a platform improves their safety standards.

› **Information gathering powers.** Ofcom can issue a notice to a platform at any time, such as requesting a risk assessment, risk mitigation measures, and/or documents from engineers about new features. This effectively ends the age of impunity and opacity in the industry.

This is similar to powers held by the Office of the eSafety Commissioner, however in the UK OSA, there are criminal sanctions available to senior directors if they fail to produce materials requested by the regulator. Australia's fine regime is substantively smaller – in 2023, X's failure to produce adequate materials was met with a fine of $610,500AUD.[3] A penalty of this size is arguably negligible for large companies with multi-billion dollar annual turnovers.

› **Strong enforcement powers.** Alongside the ability to issue criminal sanctions for failures to comply with transparency requests, Ofcom has powers around failure to adequately mitigate risks. They can issue fines of up to 10% of global annual *turnover* and have 'shutdown powers' for particularly grave breaches.



'Cloud trawling for data'
Digital illustration by Benjamin Horgan

2. THE *EU'S* ENFORCEABLE SAFETY STANDARDS

The EU's *Digital Services Act* ('DSA') is broad in comparison to Australia's *Online Safety Act*. While the proposed changes to the BOSE outlines a handful of systems or elements that may be subject to safety requirements (recommender systems, generative AI systems, user-controls etc), the DSA covers any systemic risk stemming from *design or functioning* of a platform, and any systemic risks pertaining to the *use* of the platform.

Articles 34 and 35 of the DSA are two of the more relevant clauses to understand in thinking through the breadth of the DSA's approach in an online safety domain. Essentially, these two clauses say 'as a platform, you have to make a risk assessment and you will be measured against that'. Specifically:

› **Article 34**: Requires very large online platforms to undertake risk assessments of each of these systems for risks they create regarding;
  - Illegal content;
  - Risks towards fundamental rights, such as dignity and privacy and political freedoms;
  - Risks for civic discourse and electoral processes, and public security;
  - Risks around gender-based violence, public health, children's wellbeing, and serious negative consequences to people's physical and mental well-being.

It lists some systems and elements that must be included in the risk assessments—for example, recommender systems, advertising systems, content moderation systems and data practices—but also requires all systems and elements be addressed.

› **Article 35**: Places responsibility on platforms to put in place 'reasonable, proportionate and effective mitigation measures' to address the risks identified in these risk assessments. They note a range of measures—which in Australia's BOSE proposals might be called 'reasonable steps'—including but not limited to:
  - Adapting designs, features or functionings of platforms;
  - Changing terms and conditions, or their enforcement;
  - Improving content moderation processes;
  - Adapting algorithmic systems, including recommender systems;
  - Changing advertising systems, including limiting or adjusting the presentation of ads;
  - Improving internal business processes;
  - Changing cooperation with 'trusted flaggers' (or as Australia calls them 'fact checkers');
  - Changing cooperation with other online platforms or search engines;
  - Providing end-users more information about their service;
  - Improving measures to protect children's rights, and;
  - Taking action on deep fakes and synthetic material.

Comparatively, Australia's proposed approach is a patchwork, covering only a handful of systems and largely addresses risks from illegal and harmful content, without much specification.

The DSA introduces a range of transparency obligations, and enforcement options where necessary. Transparency obligations include the development of:

- **Risk assessments**, which are transmitted directly to the regulators, but should become public after an 'extended' year with the regulators;
- **Annual, public transparency reports.** These transparency reports are heavily prescriptive under the DSA, so there is a clear 'template' of information platforms must provide and there is little room for interpretation from platforms. This also allows comparisons between providers. The first round of transparency reports released by Very Large Online Platforms under the DSA revealed a trove of information, such as the low number of moderators employed in non-English speaking markets;
- **Annual, independent audits.** Alongside the transparency reports and risk assessments—which are produced by platforms according to specifications outlined in the DSA—platforms must undertake an independent audit of the risks on their platforms and publish this independent evaluation;
- **Ad repositories,** or openly searchable databases of all ads presented on platforms including targeting options and data about advertisers;
- **Researcher access to public interest data.** Under Article 40, there is an obligation that platforms share public interest data with vetted researchers as requested, to facilitate academic and civil society analysis of risks posed.

The DSA came into force in November 2022. The key question now is how the Commission will interpret the severity of systemic risks and the efficacy of platform risk mitigations. Some suspect this interpretation will be heavily guided by the platforms' own targets. Further, some argue the Commission is unlikely to make its own assessment about systemic risks, but this may change over time.

The Commission has strong powers of redress if platforms fail to comply. They have powers to make visits, take interviews, but also to issue penalties of up to 6% of global annual turnover for failures to effectively mitigate risks, or fines of up to 1% of global annual turnover for supplying incorrect, incomplete or misleading information as part of meeting transparency obligations. In extreme cases—as a last resort, where all options available under national and EU law are exhausted—the DSA afforces service closure powers. Article 51(3) outlines that where "the infringement has not been remedied or is continuing and is causing serious harm, and that that infringement entails a criminal offence involving a threat to the life or safety of persons, to request that the competent judicial authority of its Member State order the temporary restriction of access". Note, there are no criminal sanctions under the DSA.

3. *CANADA'S* DECISION TO MOVE TOWARDS RISK-BASED, ENFORCEABLE SAFETY STANDARDS

Canada is in the process of developing an *Online Harms Act,* having presented an initial draft bill for consideration in 2021. Initially, this first Bill and early discussions regarding digital regulation were very content focussed, and mirrored much of Australia's current *Online Safety Act*. For example, the 2021 Bill addressed five categories of harmful content—CSAM, hate speech, non-consensual sharing of intimate images, terrorist content, incitement to violence—and proposed measures such as a 24 hour takedown clause. These content categories will likely sound familiar to an Australian audience.

Wide-ranging feedback from stakeholders recommended a more comprehensive, future-proofed duty of care approach. Canada may be a bit of a late mover, but this does afford the ability to observe what has worked in other jurisdictions, and adapt these lessons to emerging policy trends. Hopefully when the revised Bill is tabled and introduced, Canada will have world-leading online safety legislation.

While the revised Bill has not been publicly released yet, we are expecting it to include:

- Duties of care regarding children, coverage of a breadth of systems and elements such as algorithms and design choices;
- In terms of transparency, various reporting requirements, from risk assessments to transparency reports as well as requirements to provide researcher access to data;
- In terms of enforcement, a strong enforcement regime similar to the UK's or EU's frameworks. The thinking in policy circles points towards wanting legislation with teeth, rather than just asking platforms nicely to set their own standards and mark their own homework. We expect that enforcement will be done by empowering an independent federal regulator likely to be built from ground up, with investigative, auditing, and enforcement powers. We would imagine that the Bill would have substantial penalties for platform non-compliance as well as some sort of consumer redress mechanism.

All of these expectations need to be caveated by the understanding that the revised Bill is dynamic and under ongoing development. Further, the *Online Harms Bill* was initially developed by the Department of Canadian Heritage but is moving into the Department of Justice, which might involve other changes in direction.

ANALYSIS

Ensure **comprehensive coverage of systems and issues**

There is a strong desire for Australia's online safety framework—including the *Online Safety Act* and *Basic Online Safety Expectations*—to comprehensively cover all of a platform's elements that create systemic risk for Australian end-users. As the comparative table in *Figure 1* highlights, there are significant gaps in protection created by the existing proposals.

Introducing an overarching duty of care in the *Online Safety Act* might help to redress these gaps, as would including requirements to ensure basic safety standards are met across all systems and processes in the BOSE.

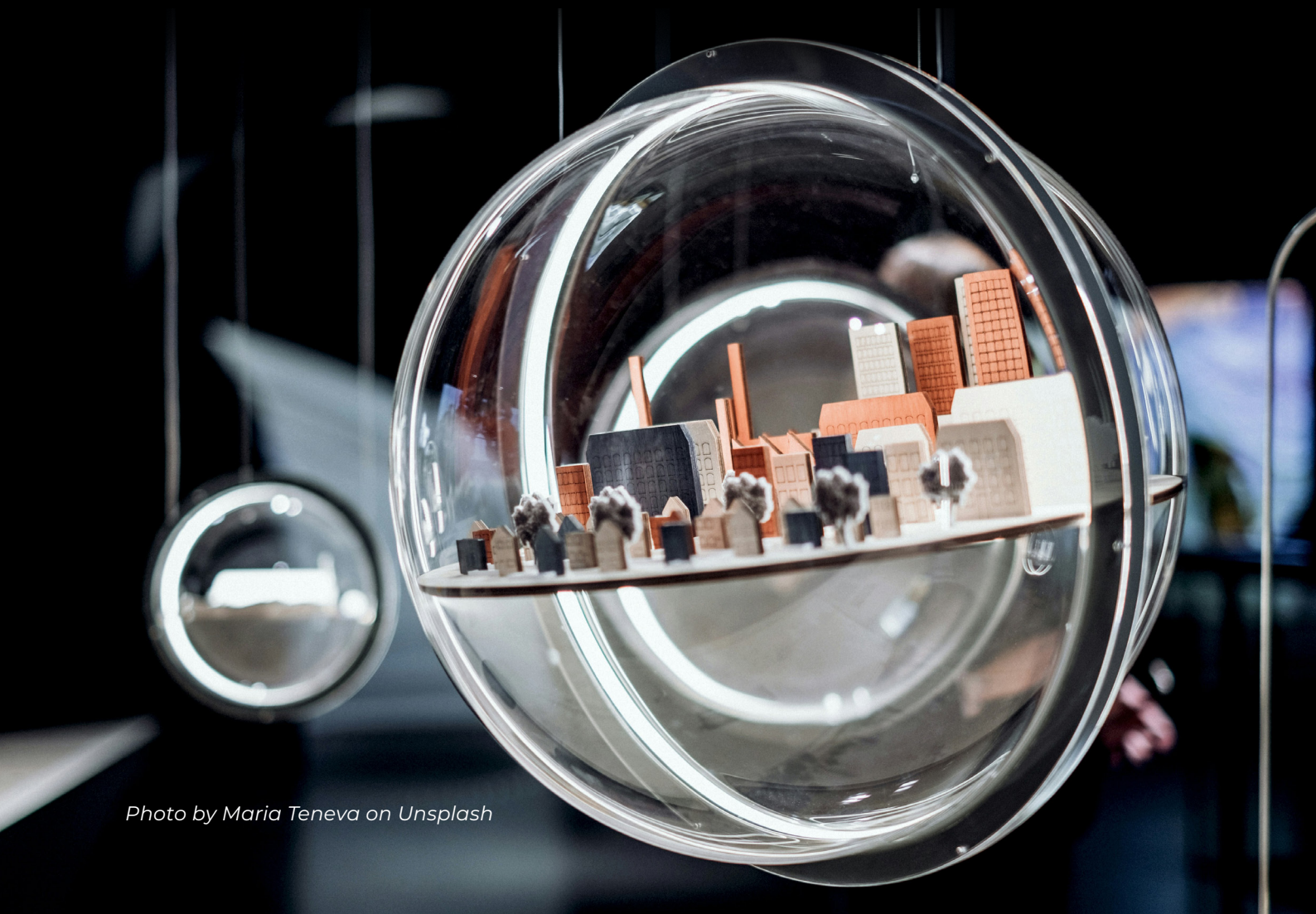


Photo by Maria Teneva on Unsplash

Figure 1: A comparative analysis between the UK's OSA, the EU's DSA and Australia's proposed protections under the BOSE scheme.

Comparison of requirements in the DSA, the UK OSA and the proposals under the BOSE

Systems 'listed' in the DSA as subject to risk assessment criteria for Very Large Online Platforms	Systems 'listed' in the UK OSA as requiring measures to ensure duties of care are met across Platforms	Systems 'listed' in the proposed BOSE as being subject to expectations regarding reasonable steps
Recital 84 outlines that services should "focus on the systems or other elements that may contribute to the risks", and lists a number of examples. Other systems and elements specifically listed across the legislation include: 1. Recommender systems 2. 'Safety by design' settings for minors 3. Dark patterns and design of interfaces 4. Advertising systems 5. Content moderation systems 6. Notice action and complaint mechanisms 7. Trusted flagger systems 8. Terms and conditions	The duties of care laid out in the Act "apply across all areas of a service, including the way it is designed, operated and used as well as content present on the service," and lists the following areas as requiring measures (non-exhaustive): 1. Regulatory compliance and risk management arrangements 2. Design of functionalities, algorithms and other features 3. Policies on terms of use 4. Policies on user access to the service or to particular content present on the service, including blocking users from accessing the service or particular content 5. Content moderation, including taking down content 6. Functionalities allowing users to control the content they encounter 7. User support measures 8. Staff policies and practices	1. Generative AI capabilities 2. Recommender systems 3. User controls 4. 'Safety by design' settings for minors (via the best interests proposal in subsection 6(2)(A)) 5. Enforcement of terms of use (14(1A)) 6. Complaints & reporting systems (14(3)) We note that some aspects around staff practices covered by the UK's OSA may be addressed by proposals to amend paragraph 6(3)(f), to add in a suggested example that services assessing whether business decisions will have a significant adverse impact on the ability of end-users to use the service in a safe manner. Further, elements of the DSA's requirements around terms and conditions regarding understandability are being explored in the *Privacy Act Review*.

There is strong support from the public for this broader approach. In January 2024, Reset.Tech commissioned YouGov to poll 1,005 Australian adults.

We found overwhelming support for including expectations regarding more systems—such as advertising systems and content moderation systems—and all systems in general *(see Figure 2)*.

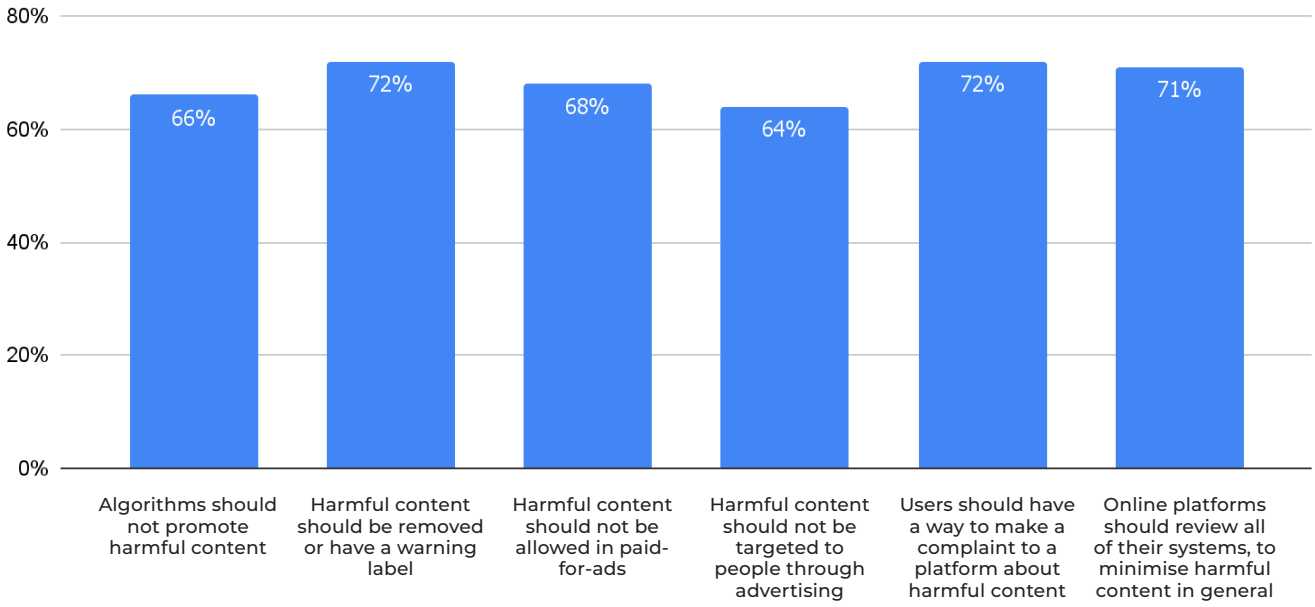


Figure 2: Responses to the question 'Which of the following do you think online safety regulations should require?' n=1,005

WHAT IS A DUTY OF CARE?

A duty of care approach draws from experience of other areas of safety regulation such as workplace health and safety which in the UK, as in Australia, is determined by a duty on the people who control and are responsible for the hazardous environment. The Carnegie Foundation outline four key aspects critical to a duty of care approach in an online environment:

1. ***The overarching obligation to exercise care in relation to user harm;***
2. ***Risk assessment process;***
3. ***Establishment of mitigating measures; and***
4. ***Ongoing assessment of the effectiveness of the measures.***[4]

In the related context of data, the Consumer Policy Research Centre has explored the idea of duty of care and best interests duty for data, and found strong support from Australians.[5]

A need for **accountability**

The UK's ability for regulators to 'hold the pen' on drafting Codes differ from the Australian experience. In Australia, industry leads on the development of Codes regarding illegal and harmful content, and then presents these Codes to be considered by the Office of the eSafety Commissioner. This process skews the bargaining power of regulators and civil society, and delivers inadequate safety standards for end-users.[6] By comparison:

- In the UK, Ofcom will be drafting Codes involving significant consultation with key stakeholders, including industry and civil society.
- The EU has a tiered approach to developing Codes. For some issues, the European Commission's role will be to encourage and support the development of voluntary Codes, which it will then oversee compliance with (as per the Australian model) but for issues presenting 'significant systemic risks' that are common across Very Large Online Platforms and Search Engines, the Commission takes a more active role. In this vision of co-regulation, the Commission is expected to lead on the Code and 'may invite the providers of very large online platforms concerned or the providers of very large online search engines concerned, and other providers of very large online platforms, of very large online search engines, of online platforms and of other intermediary services, as appropriate, as well as relevant competent authorities, civil society organisations and other relevant stakeholders, to participate in the drawing up of codes of conduct'.[7] This is a more nuanced vision of how 'co-regulation' could work in practice.

The move towards regulator drafted Codes in the UK was not accidental, and was extensively supported by academics and civil society who took learnings about the inadequacies of previous voluntary codes (including, for example, the EU's voluntary Code on Misinformation and Disinformation, see appendix one.) Polling revealed it was also extremely popular with the British public, which provided political support for the move.

Similar polling reveals extensive support across the Australian public for online codes of practice to be drafted by regulators. A poll of 1,508 Australian adults in December 2022 revealed that:

- 73% would prefer the eSafety Commissioner draft Online Safety Code(s), with only 5% preferring that the social media industry lead
- 76% would prefer the Privacy Commissioner and the Office of the Australian Information Commissioner draft Online Privacy Codes, again with only 5% preferring that the social media industry lead.[8]

A need to **enhance transparency requirements**

Both the UK's OSA and the EU's DSA have stronger transparency requirements than those proposed under the BOSE.

Where platforms are able to largely develop the structure of public transparency reports, there is a concern that platforms will still be able to control the narrative and 'spin' the contents of their reports that describe their systems and processes in unduly positive ways.[9] Under the DSA, the requirements of transparency reports are extremely prescriptive. For example, there are compulsory metrics like:

- Indicators of accuracy relating to the information provided, takedown orders received;
- Number of illegal content notices issued;
- Number of complaints received through internal complaint-handling systems;
- Human resources that the platforms dedicates to content moderation, broken down by language;
- Qualifications and linguistic expertise of the persons carrying out content moderation, as well as training and support given to such staff;
- Indicators of accuracy of content moderation, broken down by language;
- Average monthly "users" for each Member State;
- Meaningful and comprehensible information about the content moderation engaged in at the platform's own initiative.

These metrics are significantly more prescriptive than the requirements under the BOSE.

Public transparency reports created under the DSA are also supported by detailed risk assessments (transmitted directly to the European Commission), independent audits, compulsory ad repositories and importantly, researcher access to public interest data. The requirement for researcher access to public interest data places an obligation on platforms to provide vetted researchers with '*access to data … for the sole purpose of conducting research that contributes to the detection, identification and understanding of systemic risks in the Union… and to the assessment of the adequacy, efficiency and impacts of the risk mitigation measures*'.[10] This is an extremely robust measure, and advocates from the UK spoke about their 'envy' over the strength of these researcher access requirements.

Introducing requirements around researcher access should address barriers created by platforms that charge fees for access to data and APIs. This is both implied under researcher access requirements, but lawmakers can additionally include detail in the legislation that specifies what, if any, fees are to be payable by vetted researchers. It is worth noting that researcher access schemes extend beyond APIs however, and specific and bespoke data should be requestable. Affordable API access alone is insufficient.

A list of requirements regarding what counts as a vetted researcher is also necessary, and these must be limited to non-commercial researchers undertaking public interest research.

There is strong public support for increasing accountability and transparency when it comes to user safety. We polled 1,005 Australians about online regulation, and found broad support for accountability (phrased as enforcement) and transparency (phrased as oversight) *(see Figure 3)*.

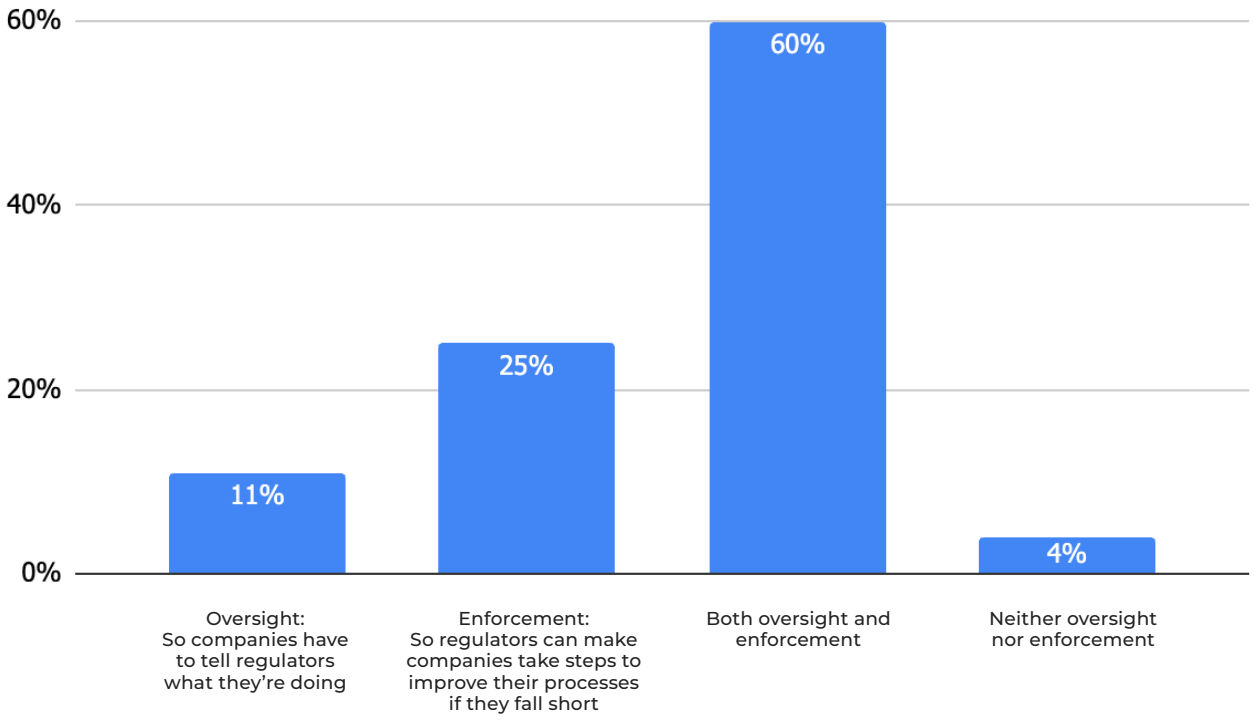


Figure 3: Responses to the question 'Thinking about online safety regulations for companies, which of these do you think should be required?' n=1,005

A need for **strong enforcement**

Both the UK's OSA and the EU's DSA afford stronger powers to compel redress, and offer significantly stronger penalties for non-compliance than Australia's *Online Safety Act*.

In part, this may reflect the European Commission's learnings around the difficulties a lack of enforcement has created when it comes to privacy and data protection under the *General Data Protection Regulation* (GDPR). The lack-lustre enforcement of the GDPR has meant that many significant privacy issues have failed to be addressed.[11] It is worth noting however, that even with these difficulties, data protections are significantly stronger in theory and in practice in the EU under the GDPR than in Australia. This comparative weakness will still hold even after the proposed reforms to the *Privacy Act* are implemented, unless significant steps are taken to strengthen the proposals.

Likewise in the UK, there was some reflection on the issues of 'enforceability' in the *Age Appropriate Design Code*, and a strong desire to see the UK's OSA significantly strengthened to ensure improvements in the digital environment for British children.

Advocates in Europe spoke about 'envy' for the UK's criminal provisions for non compliance with transparency measures. These are not available under the DSA.

There was also discussion around the level of penalties available to regulators in Australia. For example, when X (formerly Twitter) was issued with non-compliance order for failing to respond to a transparency request from the Office of the eSafety Commissioner, they were fined the maximum allowable $610,500 AUD.[12] Paying this fine would still be cheaper than resourcing the staff required to implement the meaningful safety measures in question.[13] Comparatively, under the DSA, regulators can issue fines up to 6% of a company's global annual turnover, and in the UK they can fine up to 10% of global annual turnover. Note that in Australia, regulators in the adjacent domains of consumer protection and financial services have comparable fining abilities.[14]

Beyond fines, overseas online safety regulators also have powers to:

- Compel redress, so that regulators can ensure platforms change and improve safety standards. They can issue significant fines for failures to do so;
- 'Turn off' services where the failures are significant and persistent, and attempts at engagement have failed;
- In extreme cases in the UK, issue criminal sanctions if transparency measures are not met.

There is a unique capacity for Australia to develop a world-class enforcement system. Australia already has a leading public-facing complaints mechanism under the *Online Safety Act*—a feature missing from both the EU and UK's frameworks. Public-facing complaints mechanisms should be extended, so that end-users and their representatives can make complaints to regulators regarding breaches of platform safety standards. A reform of this scale would necessarily require meaningful resourcing to the relevant regulator.

CONCLUSIONS AND RECOMMENDATIONS

A number of amendments are required to the proposals around the *Basic Online Safety Expectations* and the *Online Safety Act* to ensure that Australian end-users enjoy meaningful safety standards with equivalence to our international counterparts.

These include:

- **Ensuring that a comprehensive breadth of systems and elements are addressed under the framework.** This could include:
  - Introducing an overarching duty of care to place broad responsibilities on platforms regarding all of their systems and processes;
  - Introducing requirements for platforms to implement reasonable steps for end-user safety across all systems and elements of their platforms, including but not limited to generative AI systems, recommender systems, user-controls, enforcement of terms and conditions (and proposed by the BOSE amendments) but also content moderation systems, including fact-checking systems, advertising approval systems and advertising management systems.
- **Ensuring meaningful transparency by introducing a suite of compulsory transparency measures,** such as:
  - Risk assessments that must be submitted for review by regulators;
  - Annual transparency reports with a detailed set of requirements to ensure 'meaningful' data is reported by platforms;
  - Independent audits for larger online platforms;
  - Ad repositories for Australian ads, and;
  - Researcher access to public interest platform data.
- **Ensuring effective compliance by enhancing regulators' powers.** This includes expanding powers beyond existing transparency and 'take down' measures to;
  - Ensure the ability to compel redress and changes to platforms' systems and processes;
  - Increase penalties;
  - Enable the ability to 'turn off' services where failures are persistent and all other measures have been exhausted.
- Additionally, given Australia already has a **world-class public facing complaints mechanism** under the *Online Safety Act*—which is lacking in both the EU and UK—this mechanism should be extended to basic online safety standards, so that end-users and their representatives can make complaints to regulators regarding breaches of safety standards.

ACKNOWLEDGEMENTS

This briefing paper reflects the expertise of those who contributed to the roundtable. Attendance does not necessarily mean endorsement. Attendees included:

› Rafi Alam, *Choice*
› Aruna Anderson, *Reset.Tech Australia*
› Marilyn Bromberg, *University of Western Australia*
› Marianne Campbell, *Consumer Policy Research Centre*
› Alice Dawkins, *Reset.Tech Australia*
› Alice Drury, *Human Rights Law Centre*
› Surpiya Dwivedi, *McGill University*
› Jasmine Fardouly, *University of New South Wales*
› Rys Farthing, *Reset.Tech Australia*
› Samantha Floreani, *Digital Rights Watch*
› Hannah Jarman, *Deakin University*
› Felix Kröner, *Reset.Tech Berlin*
› John Livingstone, *UNICEF Australia*
› Noelle Martin, *University of Western Australia*
› David Mejia-Canales, *Human Rights Law Centre*
› Jacqui McKenzie, *ChildFund Australia*
› Jessie Mitchell, *The Alannah & Madeline Foundation*
› Dylan Sparks, *Reset.Tech UK*

All errors and omissions rest with Reset.Tech Australia.

Appendix

FROM VOLUNTARY CODE TO COMPREHENSIVE REGULATION: THE EUROPEAN EXPERIENCE

This timeline summarises the European experience and shows how legislators gradually responded to the shortcomings of the voluntary industry code with a more comprehensive package. Notably, requirements for data access were consistently invoked to ensure that there were mechanisms for independent assessments of what was otherwise mere platform self-reporting.

MAR 2018	APR 2018	SEP 2018	JAN 2019	MAR 2019	MAR 2019
Final report of the High Level Expert Group on Fake News and Online Disinformation	European Commission responds with a 'Code of Practice on Disinformation' which would commit online platforms and the advertising industry to provide academia with "access to platform data"	Version 1 of the Code of Practice is released	The European Commission expresses concern on the platforms' failure to benchmark and meaningfully measure progress.	The European Commission remarks platforms "didn't provide access to more granular data to assess the effectiveness of their activities to counter disinformation"	The European Commission calls for independent data access to ensure that the platforms are "not just marking their own homework"

2019-2020	SEP 2020	2020-2021	JUN 2022	NOV 2022	SEP 2023
An independent assessment by EU Media Regulators (ERGA) notes no sufficient progress was made on platform commitments under the Code.	Findings from the European Commission on the first 12 months of the Code of Practice released, noting "shortcomings mainly due to the Code's self-regulatory nature".	Draft *Digital Services Act* provisions construct a data access regime with a legal basis to force VLOPs/VLOSE to provide access to data to third Parties, including regulators, vetted researchers, and civil society organisations.	Roll-out of the 'Strengthened' Code of Practice on Disinformation.	The *Digital Services Act* enters into force, including risk mitigation duties on platforms and mandated data access for regulators, civil society organisations, and accredited researchers.	First risk mitigation reporting from platforms under the *Digital Services Act.*

ENDNOTES

1 Department of Infrastructure, Transport, Regional Development, Communications and the Arts, *Online Safety (Basic Online Safety Expectations) Amendment Determination* 2023 https://www.infrastructure.gov.au/have-your-say/online-safety-basic-online-safety-expectations-amendment-determination-2023

2 For example, Reset.Tech Australia revealed how voluntary commitments made under the *Australian Code of Practice on Disinformation and Misinformation* fail basic external scrutiny. See for example, Reset.Tech Australia 2023 *Is political content over or under moderated?* https://au.reset.tech/news/report-misinformation-in-paid-for-advertising/ and Reset.Tech Australia 2023 Is political content over or under moderated? https://au.reset.tech/news/report-is-political-content-over-or-under-moderated/. Our findings mirror the experience of the EU's voluntary disinformation code, which was deemed to be ineffective and robustly augmented with a risk assessment, risk mitigation, and data sharing regime in the *Digital Services Act* (See Appendix 1 in Reset.Tech Australia 2023 *Misinformation and disinformation regulatory frameworks* https://au.reset.tech/news/policy-briefing-misinformation-and-disinformation-regulatory-frameworks/)

3 Josh Taylor 2023 'X fined $610,500 in Australia first for failing to crack down on child sexual abuse material' *The Guardian* https://www.theguardian.com/technology/2023/oct/16/x-fined-610500-in-australia-first-for-failing-to-crack-down-on-child-sexual-abuse-material

4 Carnegie UK 2022 *Submission to the House Select Committee on Social Media* and Online Safety available at https://www.aph.gov.au/Parliamentary_Business/Committees/House/Former_Committees/Social_Media_and_Online_Safety/SocialMediaandSafety/Submissions

5 Consumer Policy Research Centre 2023 *In whose interest – Why businesses need to keep consumers safe and treat their data with care* https://cprc.org.au/in-whose-interest/

6 See for example, Reset.Tech Australia 2022 *How outdated approaches to regulation harm children and young people* https://au.reset.tech/news/how-outdated-approaches-to-regulation-harm-children-and-young-people-and-why-australia-urgently-needs-to-pivot/

6 Article 45 *Digital Services Act* https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:32022R2065

8 See for example, Reset.Tech Australia 2022 *How outdated approaches to regulation harm children and young people* https://au.reset.tech/news/how-outdated-approaches-to-regulation-harm-children-and-young-people-and-why-australia-urgently-needs-to-pivot/

9 There is already evidence that this happens in Australia, emerging from the transparency reports required under Digi's voluntary *Australian Code of Practice on Disinformation and Misinformation*. See for example, Reset.Tech Australia 2023 Misinformation in paid-for-advertising https://au.reset.tech/news/report-misinformation-in-paid-for-advertising/

10 See Article 40, *Digital Services Act* https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:32022R2065

11 Irish Council of Civil Liberties 2023, *5 years: GDPR's crisis point* https://www.iccl.ie/digital-data/iccl-2023-gdpr-report/

12 Georgie Hewson 2023 'Australia's eSafety commission fines Elon Musk's X $610,500 for failing to meet anti-child-abuse standards' *ABC* https://www.abc.net.au/news/2023-10-16/social-media-x-fined-over-gaps-in-child-abuse-prevention/102980590

13 See for example, Reset's submission to the BOSE (forthcoming)

14 Such as the ACCC for franchising violations (see ACCC nd Fines and penalties https://www.accc.gov.au/business/compliance-and-enforcement/fines-and-penalties) and ASIC for violations of ASIC administered legislation, albeit capped at $782.5million (see ASIC 2023 Fines and Penalties https://asic.gov.au/about-asic/asic-investigations-and-enforcement/fines-and-penalties/)

Reset.
AUSTRALIA

Accountability, the Online Safety Act and the Basic Online Safety Expectations:
Can safety standards be enforceable?

Cover image: Artwork created using Midjourney in response to the prompt *"imagine/ a fluffy cloud art installation at sunrise, wide aerial view, people sitting, Hasselblad H6D-100c, Sirui 50mm f/1.8 anamorphic 1.33x, --v 5.2*