

A duty of care in Australia's *Online Safety Act*

Policy briefing | April 2024



Notice-and-take-down emerged from historical approaches to regulating broadcasting, which adopted a piece-of-content-by-piece-of-content approach. *However, this approach cannot scale to meet the risks of a digital world fuelled by masses of user-generated content and leaves regulators and content moderators playing 'whack-a-mole'.*

Reset.Tech Australia
2024

Cover image: Artwork created using Midjourney in response to the prompt "imagine/ a surreal whack-a-mole scene in a terrarium hasselblad H6D-100c, Sirui 50mm f/1.8 anamorphic 1.33x, --v 5.2

CONTENTS

Introduction	4
1. Australia's online safety framework	5
2. What can we learn from duties of care in the UK?.....	8
3. A duty of care to incentivise platform action	10
Discussion	11
Recommendations.....	14
Acknowledgements	15
Appendix: 5 pillars of systemic regulation	16
Endnotes.....	18



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 21 experts from academia and civil society (including not-for-profits and research organisations) in March 2024, where we explored the opportunities and challenges of introducing a duty of care into Australia's *Online Safety Act*.¹ The event was held under the Chatham House Rule, and this briefing presents an overview of the discussion.

This discussion was timely. While online safety advocates have been calling for a duty of care in online safety regulation since 2019, there is a current policy opportunity to see this realised in Australia. In its response to the *House of Representatives Inquiry into Social Media and Online Safety* in April 2023, the Government announced its intention to bring forward the statutory review of the *Online Safety Act* from 2025 to 2024.² In March 2024, the Department of Infrastructure, Transport, Regional Development, Communications and the Arts released the Terms of Reference for this review, which includes explicit proposals to consider:

Whether the regulatory arrangements, tools and powers available to the Commissioner should be amended and/or simplified, including through consideration of: a. the introduction of a duty of care requirement towards users (similar to the United Kingdom's Online Safety Act 2023 or the primary duty of care under Australia's work health and safety legislation) and how this may interact with existing elements of the Act.³

The review is being led by Delia Rickard, former Deputy Chair of the Australian Competition and Consumer Commission, and must report back by October 31st 2024. While the review is not open for public input yet, there will be an Issues Paper accompanied by a call for public submissions in the first half of 2024. The purpose of this discussion was to prepare civil society for this debate and to help inform the review team as they develop their issues paper.

This memo summarises the discussions at the policy roundtable. The event included three provocations, which are summarised below, as well as the broader discussion.

1. AUSTRALIA'S ONLINE SAFETY FRAMEWORK

The path to Australia's *Online Safety Act* has left us with strong content-focused regulations but weak systemic regulations

Australia's current online safety framework stems from progressive changes from the early *Enhancing Online Safety for Children Act 2015*.⁴ The Act established the Office of the Children's eSafety Commissioner and set in place rules about the removal of content that contained bullying material aimed at children. It included the introduction of a complaints service for children and families about specific content, and gave the Children's eSafety Commissioner powers around community education. This was world-leading at the time. In 2017, the Act was amended to expand the focus and was renamed *Enhancing Online Safety Act 2015*⁵ (and the regulator became the Office of the eSafety Commissioner). The amendments expanded the notice-and-take-down powers to content that included non-consensual intimate images and expanded the remit of the Commissioner to the safety of all Australians. In 2021, this was replaced with the *Online Safety Act 2021*,⁶ which again added to the list of content that was eligible for the public complaints service and subject to notice-and-take-down powers, now expanded to child-bullying content, non-consensual intimate images and adult cyber-abuse content. Notice-and-take-down powers were also extended to illegal materials (child sexual abuse, pro-terror and abhorrent violent materials).

These content-focused approaches, realised through notice-and-take-down powers, are mandatory and enforceable. There are clear expectations that if a regulator issues a take-down notice to a platform, they must take down the individual piece of content within 24 or 48 hours or risk a fine applied for each piece of content.

However, focusing on individual pieces of content will only get you so far, and it leaves regulators playing 'whack-a-mole', where all they can do is request that successive pieces of individual content be taken down. To help address this shortcoming, the *Online Safety Act 2021* also included an additional section of 'rules' that was intended to lead to systemic upstream change, called the Basic Online Safety Expectations.⁷ These expectations are set by Ministerial decree and lay out the upstream,⁸ system-wide steps digital platforms are expected to take to more comprehensively ensure the safety of users. At the moment, these include an expectation that platforms will take largely unspecified 'reasonable steps' to ensure the safety of end-users. They also provide some examples of reasonable steps that might be taken.

The Office of the eSafety Commissioner has powers to request information about what these reasonable steps are but not to enforce or demand any particular steps themselves. These expectations are, therefore, largely voluntary and unenforceable.⁹ We have documented at length how ineffectively these have been implemented, especially through the use of industry-drafted codes to establish reasonable steps,¹⁰ including noting where they appear to have actively reduced minimum protections as laid out in the legislation itself.¹¹

There are no regulatory powers in Australia to demand specific safety standards or require particular 'reasonable steps' even where these are best practice. The path Australia has taken in legislative development has left us with strong regulations regarding content but very weak regulatory oversight on the risk-producing systems themselves.

A systemic framework is both needed and possible

The review of the *Online Safety Act* presents the opportunity to develop a comprehensive and enforceable regulatory framework to address digital risks. We believe this includes five key elements (see also Appendix 1):

1. An overarching duty of care that would place broad obligations on platforms to focus on their systems and mitigate against harms before they happen.
2. Requirements for platforms to assess *all their systems and elements* for serious risks they may pose.¹² This would incentivise digital platforms to identify systemic risks and help realise their duty of care. Risk assessments could be focused on addressing key priority areas, including systems and content. As one example, we would explicitly recommend an assessment around children's best interests¹³ to ensure their rights are advanced and to help create harmonisation with the *Privacy Act* (see Appendix 1 for other examples).
3. Requirements for risk mitigation measures. As a corollary of risk assessments, platforms must be required to implement reasonable steps to mitigate each risk identified.
4. An effective framework for public transparency. This should include a suit of measures, such as:
 - i. Annual risk assessments
 - ii. Annual public transparency reports, which need to be heavily prescriptive with detailed requirements¹⁴
 - iii. Annual independent audits of risk assessments and transparency reports
 - iv. Ad repositories that document all paid-for advertising and details in searchable ways
 - v. Researcher access to public interest data.
5. Strong enforcement powers. To meaningfully drive change, regulations need to be enforceable, and regulators must be empowered and resourced to:
 - i. Compel redress and changes to platforms' systems and processes rather than just the ability to compel transparency or take-down in a timely manner
 - ii. Issue penalties that match the scale of the global profits of digital platforms
 - iii. Have powers to 'turn off' services where failures are persistent and all other measures have been exhausted
 - iv. Enhance the public-facing complaints mechanism to include complaints from individuals and consumer groups regarding systemic risks and breaches of their duty of care
 - v. Have strong investigative and information-gathering powers
 - vi. Have effective notice-and-take-down powers.

2. WHAT CAN WE LEARN FROM DUTIES OF CARE IN THE UK?

The UK passed the *Online Safety Act* (UK OSA) in late 2023.¹⁵ It includes multiple duties of care (see Figure 1).

A duty of care approach is systemic

A duty of care approach substantially differs from a notice-and-take-down approach to digital regulation. Notice-and-take-down emerged from historical approaches to regulating broadcasting, which adopted a piece-of-content-by-piece-of-content approach. However, this approach cannot scale to meet the risks of a digital world fuelled by masses of user-generated content and leaves regulators and content moderators playing 'whack-a-mole'.

A duty of care approach is a way to implement systemic regulation that moves the focus beyond the content layer to the underlying systems – the environment where content is created, shared and promoted. The design of these underlying systems is entirely within a platform's control (less so where content is generated by users). Focusing regulation on systems and processes creates a situation where platforms are required to consider whether there is a risk of harm to users arising from their technical systems, design and business models while allowing users to express themselves.

Focusing on design and operation is important because platforms are not entirely neutral, passive transmitters when it comes to content. Intentionally or not, their choice of architecture has an impact on content. This includes the role of recommender and content-moderation systems, for example, and how engagement features are designed to create social pressures or anonymous accounts. Duty of care is a way to implement systemic regulation that can address these types of risks.

'A duty of care' is more systemic than 'duties of care'

In the UK, initial academic proposals for a duty of care¹⁶ were eventually implemented as multiple 'duties of care' (see Figure 1). Implementing multiple duties of care requires distinguishing between different types of content—such as criminal content, content that is harmful to children and, for larger platforms, content harmful to adults—and associating specific duties to each type of content. For example, there is a duty to do a risk assessment and a duty to mitigate identified risks for each content type, with additional specificity about what this means for each content type. This includes specificity about minimising the amount of time particular types of content are available online and minimising the chance of people encountering particular types of content—illegal content, for example. In a sense, this approach 'calibrates' duties to match particular harms emerging from particular types of content.

This approach has some advantages. It allows for a very long and detailed identification of the types of content that regulators can address, and this helps distinguish between 'bad content' and 'really bad content'.

However, it introduces a number of challenges. Specifically:

- › It introduces incredible complexity to the regulatory framework. For example, to develop the first set of *online safety codes of practice around illegal content*, Ofcom (the regulator) needed to split the consultation document into six volumes with 16 annexes (totalling over 1500 pages) to cover the necessary details.¹⁷
- › More importantly, it moves the regulation away from a focus on the systems and back into specifying particular types of content. While the UK OSA¹⁸ is based on systemic regulation—with risk assessment and mitigation at the core of the duties of companies—the overriding and complex focus on content overlaid on top skews the focus of compliance towards a content-first rather than a systems-first approach.
- › This introduces a very particular tension. A systemic approach acknowledges that systems are developed and business decisions are made about them before they are populated with content. Requiring platforms to think about risk assessing their systems only after they are ‘populated’ with particular sorts of content reduces the broader efficacy of the approach. For example, it is not clear how online harms arising from overarching abusive designs that do not fall into a specific category of content—such as dark patterns, extended use designs or manipulative choice architectures—could be addressed under this system.
- › It also introduces particular difficulties when dealing with illegal content. Under criminal law, there are questions about *mens rea* (or intent) and the mental state of people who commit crimes, which can only be determined after the fact. A platform’s systemic obligations now potentially sit downstream of this, which makes it very complicated to take an upstream systemic approach. There is a circularity here, and it is unclear how this will play out in the UK.¹⁹


 <p>All user-to-user systems have duties regarding:</p> <ul style="list-style-type: none"> › Illegal content risk assessments › Illegal content › Content reporting › Complaints procedures › Freedom of expression and privacy › Record keeping and review. 	<p>All services likely to be accessed by children have duties regarding:</p> <ul style="list-style-type: none"> › Children’s risk assessments › Protecting children’s online safety. 	<p>The largest online services also have additional duties regarding:</p> <ul style="list-style-type: none"> › Adult risk assessment duties › Duties to protect adults’ online safety › Duties to protect content of democratic importance › Duties to protect journalistic content.
---	--	--

Figure 1: Overview of the duties of care present in the UK’s Online Safety Act²⁰

3. A DUTY OF CARE TO INCENTIVISE PLATFORM ACTION

A mandatory duty of care can be considered a layer of insurance that sits around platforms' decision-making behaviours and processes and incentivises safety. A duty of care will function to incentivise platform action against risks in multiple ways. For example:

1. **A duty of care incentivises platforms' investment in risk assessments and proactive mitigations** and encourages cross-platform collaborations. Placing a duty of care on platforms makes it clear that not knowing about reasonably foreseeable risks is not a viable defence, and there are enforceable expectations that platforms will clearly identify risks in their services. This broader obligation is critical. If mandatory risk assessment processes were implemented without it, risk assessments would become ineffective activities where platforms would, in effect, just be waiting for a regulator to tell them what they missed. Keeping in mind that platforms have significantly more resources at their disposal than regulators, this does not make sense. A duty of care requires platforms to deploy their resources to engage in proactive risk identification and management.
2. **A duty of care incentivises internal curiosity about risks in situations that often lack diversity.** One of the reasons platforms struggle to address risks, especially risks to young people, is because there is a lack of diversity and lack of exposure to risk among their teams. The social software workforce is largely composed of young, educated and privileged people. During the time in their careers when they develop software, the majority will not have had children yet, creating a perfect bubble of ignorance around harms that affect kids. People building social software bring their own perspectives and experiences to design problems, and, in many cases, those experiences are narrow. Placing a duty of care on platforms pushes them to go beyond the core experiences of their team and have more curiosity about the broader risk space they operate in.
3. **Ensure that regulatory incentives on platforms are always 'up to date'** and spurring action against new and emerging risks. Regulation needs to be broad and flexible to ensure that it can cover the breadth of evolving risks in the online world in real time. Online spaces are dynamic, and the way the public interacts with them is fluid; if we focus regulations exclusively on enumerating prohibitive types of content or behaviours, we end up creating a situation where there will always be a 'legal lag' where regulation routinely falls behind emerging risks. Narrow regulation leaves regulators chasing behaviour and harms, and does not encourage platforms to respond to emerging risks with the urgency required.

DISCUSSION

The discussion centred around four key themes. They were:

01. Duty of care and enforcement

A duty of care would need to be enforceable and enforced to improve user safety. Enforcement can take a number of forms.

Direct remedy

Australia has some globally unique possibilities here. We already have a public-facing complaints mechanism under the existing *Online Safety Act* that allows people who have been harmed by content to make a direct complaint to a regulator to seek redress. This could be expanded to allow users who have been harmed as a result of a failure of a platform's broader duty of care to also make a complaint and seek redress. This would contrast with the EU's process, where, by and large, end-users cannot make complaints directly to regulators. However, regulators may consider individual users' complaints to platforms and data about this as part of their evidence-gathering process to inform enforcement action. This mirrors the UK experience. Under the UK OSA,²¹ there are no individual causes of action. However Ofcom, the regulator, has the power to assess whether companies are complying with their duties under the Act, and in determining this, they can take into account user complaints and evidence from users. While there was some discussion in the UK about the possibility of a public-facing, ombuds-style service becoming available, the UK OSA is not set up like that.

A duty of care alone is no magic bullet, and a public-facing complaints system would be necessary to ensure a duty resulted in tangible, meaningful recourse for people who have experienced online harm as a result of a platform failure. British Columbia provides an emerging alternative model to Australia's. A narrow, content-focused act—the *Intimate Images Protection Act*²²— has been introduced, focusing on the non-consensual sharing of intimate imagery. Whereas under the Australian system, a complaint would need to be assessed by the regulator before a take-down notice can be issued, the British Columbian proposal streamlines this to require platforms to offer a 'speedy complaints-resolution process' where affected users can seek redress directly from a platform, and a platform must respond. This streamlined approach, albeit implemented for a narrow content focus, could be reviewed for appropriateness in Australia.

Regulator action

Ultimately, however, a duty of care model would require a very different enforcement model than the current notice-and-take-down system. It would largely rest on regulators understanding the quality and efficacy of risk assessment. In the UK, for example, Ofcom has a multi-stage enforcement process that rests on the ability to assess whether the companies have carried out an effective risk assessment and put in appropriate mitigating measures. However, it is still in the early days of assessing the broader impact of this approach, and Ofcom has started a period of consultation on how they will enforce the Act. However, they have been signalling that they expect a step change in platforms' performance.

02. Duty of care and transparency

Alongside enforcement, transparency measures are also needed to ensure that a duty of care is effectively realised. Without transparency, many of the risks that platforms should be required to mitigate remain invisible, and the efficacy of their responses is untested. A broad approach to transparency is required (see Figure 2).²³

Transparency cannot be left to whistleblowers alone. While whistleblowers have incredibly important information to share, it cannot be left solely to brave individuals to make necessary disclosures. For example, recent whistleblower Arturo Béjar shared vitally important information about the lack of safeguarding measures for children on Instagram and Facebook²⁴ despite the platforms' public declarations regarding child safety measures.²⁵ Documents Béjar released highlighted that Meta knew that an eighth of 13 to 15-year-olds had experienced an unwanted sexual advance on Instagram within a week but had no way to report it.²⁶ Merely hoping for more whistleblower testimonies on this sort of vital information on safety processes is not sufficient; platforms should be required to be transparent about these sorts of systemic risks.

Currently, there is not enough diverse evidence to effectively interrogate claims made by platforms. Researcher access requirements are one mechanism that is working to provide transparent data access. For example, under the EU's *Digital Services Act* (DSA),²⁷ there is a requirement for platforms to share data with vetted researchers who are investigating systemic risks. These requirements extend beyond universities and include vetted NGOs. This helps to provide the evidence necessary to interrogate platforms' claims. It can also broaden the scope and perspectives of investigations, and NGOs can come from and represent affected diverse communities. The more researchers there are asking different questions, the more robust understanding one can generate about the scale and breadth of online risks.

5 potential transparency measures for Australia



Derived from the *Digital Services Act*,²⁸ Reset.Tech Australia recommends five broad measures of transparency:

1. Annual risk assessments. These risk assessments serve many purposes, including a transparency purpose. While they are transmitted directly to the regulators for discussion and analysis, slimmed-down summaries can also become public over time.
2. Annual public transparency reports. Heavily prescriptive transparency reports can be required to provide a clear 'template' of information. Platforms have produced their first round of transparency reports under the DSA, which revealed a trove of information, such as the low number of moderators employed in non-English speaking markets, for example.²⁹
3. Annual independent audits. Alongside the risk assessments and transparency reports drafted by platforms, they should commission and publish an independent audit of the risks on their services.
4. Ad repositories. These are openly searchable databases of all ads presented on platforms, including targeting options and data about advertisers. We would recommend that these also include meta-data about ads rejected for increased full transparency.
5. Researcher access to public interest data. Australia could introduce research access requirements similar to article 40 of the DSA.

Figure 2: Five potential transparency measures that could be implemented in Australia

03. Duty of care as a culture-shifting incentive

A lack of diversity across the board—from the composition of social media software teams to researchers investigating platforms' failings—harms the introduction of broad safeguards. A duty of care can incentivise a broader range of risks to be identified and analysed by platforms and 'civil society watchdogs', resulting in stronger safeguards protecting a broader range of communities.

04. The space for advocacy in the process

The third sector is outgunned by industry but plays an important role in advocating for regulations that improve user safety. Platforms are well-resourced and incentivised to advocate for reduced regulatory oversight, so civil society needs to be strategic and well organised in order to provide an effective counterbalance.

The UK experience could be instructive. In the UK, civil society played a significant role in clearly documenting and demonstrating the harms occurring on platforms, which generated strong political support for action. In this space, the children's lobby is always influential, but broader coalitions can be found. For example, in the UK, consumer groups mobilised around online fraud and financial scams and worked with financial sector organisations to create a broad public and private advocacy movement.

Mental health and suicide prevention charities were also instrumental in documenting the breadth of support for reform in the UK. The UK ended up with a large network that spanned a number of different interests and groups, all calling for an effective UK OSA. It is worth noting that a lot of misinformation- and disinformation-focused organisations were also involved in the UK debates, although misinformation was ultimately removed from the final UK OSA because the political debate at the time was held captive by the culture wars. This debate failed to highlight how systemic approaches and transparency actually amplify freedom of speech by making platforms' actions and effects on speech visible. It shines daylight onto currently invisible practices undertaken by platforms that may limit or promote certain types of speech.

Creating a broad coalition requires extensive groundwork and collaboration, but, ultimately, may be crucial as we campaign for a more systemic-focused *Online Safety Act* in Australia.

The EU approach to responsabilising platforms in the absence of a duty of care



A duty of care is a legal concept with a long history in the UK and Australia. It creates a responsibility on entities and individuals to provide a reasonable standard of care to avoid reckless or avoidable harm to others, with consideration to the foreseeability of the harm and relationship between the person harmed and the entity or individual implicated. In the UK and Australia, duty of care is a well developed concept in common law and statute.

In the EU, some member states have comparable concepts of duty of care (such as France) but these are newer and not consistent across the union.

In the absence of a comprehensive duty of care concept, the EU's DSA places a responsibility on platforms to keep users safe in the most comprehensive sense, by referring to charter rights, and enumerating the types of 'safety and freedoms' platforms must provide. Specifically, it creates obligations on platforms to address risks posed by:

- › Illegal content
- › Negative effects for the exercise of fundamental rights such as dignity and privacy and political freedoms as outlined in the European Charter
- › Negative effects on civic discourse and electoral processes, and public security
- › Negative effects on gender-based violence, public health, children's wellbeing, and serious negative consequences to people's physical and mental well-being
- › Australia does not currently have a comparable charter of rights.

RECOMMENDATIONS

Three recommendations can be drawn from this discussion:

- 1. The review of the *Online Safety Act* should seriously explore the introduction of a duty of care** into Australia's online safety frameworks. A singular duty of care, rather than multiple content-focused duties of care, arguably provides the most comprehensive safeguarding capacity and could introduce strong incentives for platforms to change behaviour.
- 2. A duty of care would need to be matched with requirements to assess and mitigate risks, backed by strong enforcement and transparency measures.** A duty of care is not a magic bullet. Other essential ingredients include risk assessment and mitigation processes, effective transparency measures and strong enforcement powers. A duty of care has the potential to be a central and powerful pillar for driving systemic online safety changes (see also Appendix 1).
- 3. Advocacy and research organisations should connect and collaborate as the *Online Safety Act* is reviewed.** The experience of the UK suggests that creating a broad coalition of organisations and expertise, including lived experience, and advocating for effective change can drive regulatory reforms in the face of well-resourced industry opposition.

ACKNOWLEDGEMENTS

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

This includes:

- › Aruna Anderson, Reset.Tech Australia
- › Professor Anna Bunn, Curtin University
- › Alice Drury, Human Rights Law Centre
- › Sarah Davies, the Alannah & Madeline Foundation
- › Alice Dawkins, Reset.Tech Australia
- › Dr Rys Farthing, Reset.Tech Australia
- › Chandni Gupta, Consumer Policy Research Centre
- › Frances Haugen, Beyond the Screen
- › John Livingstone, UNICEF Australia
- › James McDougall, Australian Child Rights Taskforce
- › Jacqui McKenzie, ChildFund Australia
- › David Mejia-Canales, Human Rights Law Centre
- › Dr Jessie Mitchell, the Alannah & Madeline Foundation
- › Dr Maeve Walsh, *Online Safety Act* Network
- › Professor Lorna Woods, University of Essex (via video)

All errors and omissions rest with Reset.Tech Australia.



AUSTRALIAN
CHILD RIGHTS
TASKFORCE



alannah & madeline
foundation

ChildFund
Australia

CPRC
Fairer markets for Australians

Human
Rights
Law
Centre

Appendix:

5 PILLARS OF SYSTEMIC REGULATION

1. A duty of care shaping platforms' actions	2. Requirements for risk assessments	3. Requirements for risk mitigations	4. Requirements for transparency measures	5. Requirements for accountability measures
<p>An overarching duty of care would place broad obligations on platforms to ensure user safety in systemic ways. Specific responsibilities could be enumerated by focusing requirements for risk assessments.</p> <p>The UK OSA³⁰ introduces duties of care, and draft Canadian legislation³¹ introduces duties on services; however, both are pluralised, which reduces the systemic focus.</p> <p>The EU's DSA³² regulations have similar systemic obligations but are phrased as responsibilities to address particular risks, specifically risks posed by:</p> <ul style="list-style-type: none"> • Illegal content • Negative effects for the exercise of fundamental rights, such as dignity and privacy and political freedoms, as outlined in the European Charter • Negative effects on civic discourse and electoral processes and public security • Negative effects on gender-based violence, public health, children's wellbeing and serious negative consequences to people's physical and mental wellbeing.³³ 	<p>Requirements for platforms to assess <i>all their</i> systems and elements for risks would incentivise systemic change and help platforms realise their duty of care.</p> <p>Risk assessments could be focused on addressing the following:</p> <ul style="list-style-type: none"> • Illegal materials (such as class 1A & 1B materials as already included in the <i>Online Safety Act</i>) • Harmful materials (such as image-based abuse, abhorrent violent material, cyberbullying of children and abuse of Australian adults as already included in the <i>Online Safety Act</i>) • Online scams • Risks to electoral processes and public security • Risks to human rights, such as political freedoms, hate speech and violence or serious harm to individuals • Risks to gender-based violence, children's best interests, public health and the environment. <p>These risk assessments should go to regulators, and summaries could be made public after a period of time.</p> <p>Both the EU's DSA and the UK OSA require risk assessments.</p>	<p>As a corollary of risk assessments, platforms must be required to implement reasonable steps to mitigate against each risk identified.</p> <p>These measures must be included in the assessments sent to regulators.</p> <p>Both the EU's DSA and the UK OSA require risk mitigations against risks identified in assessments. Canada's Online Harms Bill also places obligations on platforms to mitigate risks aligning with their duties.</p>	<p>Five different measures could be introduced to enhance public transparency:</p> <ul style="list-style-type: none"> • Annual risk assessments • Annual public transparency reports, which are heavily prescriptive • Annual independent audits of risk assessments and transparency reports • Ad repositories. Openly searchable databases of all ads and meta-data about ads rejected • Researcher access to public interest data and requirements that vetted researchers can access public interest data. <p>These need to exist alongside strong investigative powers for regulators.</p> <p>The EU's DSA requires this sort of public transparency regime.</p>	<p>To meaningfully drive change, regulations need to be enforceable. Specifically, regulators must be empowered and resourced to:</p> <ul style="list-style-type: none"> • Compel redress and changes to platforms' systems and elements rather than just compel transparency or take-down • Issue penalties that match the scale of global profits of digital platforms • Have powers to 'turn off' services where failures are persistent and all other measures have been exhausted • Enhance the public-facing complaints mechanism to include complaints from individuals and consumer groups regarding systemic risks and breaches of their duty of care • Have strong investigative and information-gathering powers • Effective notice-and-take-down powers. <p>The European Commission, the UK's Ofcom and the new Canadian Digital Safety Commissioner have combinations of these enforcement powers. The UK goes even further and includes criminal sanctions for executives regarding transparency.³⁴</p>

ENDNOTES

- 1 Commonwealth of Australia 2021 *Online Safety Act*, <https://www.legislation.gov.au/C2021A00076/latest/text>.
- 2 Australian Government 2023 *Australian Government response to the House of Representatives Select Committee on Social Media and Online Safety report* <https://www.infrastructure.gov.au/sites/default/files/documents/australian-gov-response-to-house-of-reps-select-committee-on-social-media-and-online-safety-report-march2023.pdf>
- 3 Department of Infrastructure, Transport, Regional Development, Communications and the Arts 2023 *Terms of Reference – Statutory Review of the Online Safety Act 2021* <https://www.infrastructure.gov.au/sites/default/files/documents/tor-statutory-review-online-safety-act-2021-8Feb.pdf>
- 4 Commonwealth of Australia 2015 *Enhancing Online Safety for Children Act*, http://www5.austlii.edu.au/au/legis/cth/num_act/eosfca2015321/
- 5 Commonwealth of Australia 2017 *Enhancing Online Safety for Children Amendment Act* <https://www.legislation.gov.au/C2017A00051/asmade/text>
- 6 Commonwealth of Australia, *Online Safety Act 2021* <https://www.legislation.gov.au/C2021A00076/latest/text>
- 7 Commonwealth of Australia 2022 *Online Safety (Basic Online Safety Expectations) Determination* <https://www.legislation.gov.au/F2022L00062/asmade/text>
- 8 Upstream here describes an approach to solutions that sit 'upstream' of the harms, i.e. implement changes to prevent a harm occurring 'downstream'
- 9 Reset.Tech Australia 2024 *Can safety standards be enforceable?* <https://au.reset.tech/news/briefing-can-safety-standards-be-enforceable/>
- 10 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/>
- 11 For example, the Basic Online Safety Expectations states an expectation that 'if a service or a component of a service (such as an online app or game) is targeted at, or being used by, children ... ensuring that the default privacy and safety settings of the children's service are robust and set to the most restrictive level,' and presumably rests on the *Online Safety Act's* definition of a child which is 'an individual who has not reached 18 years.' The Online Safety Code that emerged from an industry writing activity to operationalise these requirements reduced the age at which children were protected to 16 (See for example Australian Child Rights Taskforce 2023 *Letter to the eSafety Commissioner* <https://childrightstaskforce.org.au/wp-content/uploads/2023/01/Online-Safety-Codes--ACRT-letter-to-eSafety.pdf>)
- 12 For a discussion around why protections must be extended to all systems and elements, see Reset.Tech Australia 2024 *Not Just Algorithms*. <https://au.reset.tech/news/report-not-just-algorithms/>
- 13 Drawing on the United Nations Convention on the Rights of the Child, United Nations 2013 *Convention on the Rights of the Child* https://www2.ohchr.org/english/bodies/crc/docs/gc/crc_c_gc_14_eng.pdf. See also emerging work around the Best Interests Principle in the digital world at Digital Futures Commission 2024 *The best interests of the child in the digital environment* <https://www.digital-futures-for-children.net/best-interests>
- 14 For a discussion around the need for prescriptive transparency reports, see Reset.Tech Australia 2024 *Regulating for Transparency: Transparency Reports in Australia* <https://au.reset.tech/news/briefing-transparency-reports-in-australia/>
- 15 HM Government 2023 *Online Safety Act*. <https://www.legislation.gov.uk/ukpga/2023/50/enacted>
- 16 Lorna Woods and Will Perrin 2019 *Online harm reduction – a statutory duty of care and regulator* *Carnegie Trust* https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4003986. See also Lorna Woods 2019 'The duty of care in the Online Harms White Paper' *Journal of Media Law* https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4003986
- 17 Ofcom 2023 *Consultation: Protecting people from illegal harms online* <https://www.ofcom.org.uk/consultations-and-statements/category-1/protecting-people-from-illegal-content-online>
- 18 Chapter 2, HM Government 2023 *Online Safety Act*. <https://www.legislation.gov.uk/ukpga/2023/50/enacted>
- 19 There is a challenge here similar to that found in consumer law. Historical common law consumer protection remedies focus on the transaction, then the intent and only finally, the harm caused. It is not a systemic or preventative approach.
- 20 HM Government 2023 *Online Safety Act*. <https://www.legislation.gov.uk/ukpga/2023/50/enacted>
- 21 HM Government 2023 *Online Safety Act*. <https://www.legislation.gov.uk/ukpga/2023/50/enacted>
- 22 British Columbia Government 2023 *Intimate Images Protection Act* https://www.leg.bc.ca/content/data%20-%20ldp/Pages/42nd4th/3rd_read/PDF/gov12-3.pdf
- 23 For example, see Reset.Tech Australia 2024, *Regulating for Transparency: Transparency Reports in Australia* <https://au.reset.tech/news/briefing-transparency-reports-in-australia/>
- 24 Arturo Béjar 2023 *US Subcommittee on Privacy, Technology, and the Law, November 7, 2023* <https://www.judiciary.senate.gov/imo/media/doc/2023-11-07--testimony--bejar.pdf>
- 25 Barbara Ortutay 2023 *Meta sued over claims Facebook and Instagram are harming kids' mental health* *Sydney Morning Herald* <https://www.smh.com.au/business/companies/meta-sued-over-claims-facebook-and-instagram-harm-children-s-mental-health-20231025-p5ees1.html>
- 26 Zoe Kleinman, Tom Gerken & Liv McMahon 2023 'I blew the whistle on Meta, now I won't work again' *BBC* <https://www.bbc.com/news/technology-67343550>
- 27 EU 2022 *Digital Services Act Regulation*. <https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:32022R2065>
- 28 Ibid.
- 29 For example, see European Commission 2023 *DSA Transparency Database* <https://transparency.dsa.ec.europa.eu/>
- 30 HM Government 2023 *Online Safety Act 2023* <https://www.legislation.gov.uk/ukpga/2023/50/enacted>
- 31 Canada 2024 *Online Harms Bill 2024* <https://www.parl.ca/LegisInfo/en/bill/44-1/c-63>
- 32 EU 2022 *Digital Services Act* <https://eur-lex.europa.eu/eli/reg/2022/2065/oj>
- 33 Article 34, EU 2022 *Digital Services Act* <https://eur-lex.europa.eu/eli/reg/2022/2065/oj>
- 34 UK Department for Science, Innovation & Technology 2024 *Online Safety Act: new criminal offences circular* [https://www.gov.uk/government/publications/online-safety-act-new-criminal-offences-circular/online-safety-act-new-criminal-offences-circular](https://www.gov.uk/government/publications/online-safety-act-new-criminal-offences-circular/online-safety-act-new-criminal-offences-circular/online-safety-act-new-criminal-offences-circular)



Cover image: Artwork created using Midjourney in response to the prompt "imagine/ a surreal whack-a-mole scene in a terrarium hasselblad H6D-100c, Sirui 50mm f/1.8 anamorphic 1.33x, --v 5.2