Reset.
AUSTRALIA

Response to the Statutory Review of the *Online Safety Act 2021*

Reset.Tech Australia
June 2024

Contents

About Reset.Tech Australia & this submission

We welcome the opportunity to respond to the Department of Infrastructure, Transport, Regional Development, Communications and the Arts' *Statutory Review of the Online Safety Act: Issues Paper*. Reset.Tech Australia is an Australian policy development and research organisation. We specialise in independent and original research into the social impacts of technology. We are the Australian affiliate of Reset.Tech, a global initiative working to counter digital harms and threats. Our networked structure opens up strong comparative possibilities with other jurisdictions, such as in the EU, where the *Digital Services Act* is in operation, the UK, which has just passed an *Online Safety Act* and Canada, where an Online Harms Bill has been introduced to Parliament.

In this submission, we outline an overall approach to online safety that has the capacity to transform the existing approach into a more comprehensive, preventative and systemic approach. We then respond to the specific proposals put forward in the *Issues Paper*. Our proposals are informed by original research undertaken in Australia, as well as comparative policy analysis, where we draw on examples of best practice policy emerging around the world.

This submission sits alongside submissions Reset.Tech are supporting to enable children and young people's input into the discussions. We have supported Y4OS, a youth-led initiative, to reflect some perspectives of 18-25 year old Australians. In addition, we are working with the Australian Youth Affairs Council to share some perspectives of 13-17 year old Australians at a later date.

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.

1. A systemic, comprehensive, preventative focus in online safety regulations

Australia has a proud history as a 'first-mover' and innovator on digital platform regulation. Australia was the first country to legislate for online safety and introduce an online safety commissioner,[1] but has also been a strong first mover in other areas of digital regulation, such as legislating for negotiations between digital platforms and news providers.[2] Despite Australia's early mover status, the evolving nature of online risks and harms has meant that our regulatory framework has struggled to keep pace and Australians are increasingly exposed to digital risks (See Figure 1).

But we are not alone in facing the new scale and nature of these risks. Powerful new schemes are now in place in the UK[3] and the EU,[4] and a new online safety framework is under debate in Canada.[5] These jurisdictions have drawn upon the innovations and examples of Australian policy innovation to introduce comprehensive, preventative, and muscular regulatory models. These models encourage platform conduct that ensures user safety is baked into digital products and is more commensurate with public expectations for digital regulation more broadly.

Drawing on international models of regulation, Reset.Tech has identified five building blocks necessary in a regulatory framework. Below, we expand on each of these building blocks describing how they could be implemented in Australia's *Online Safety Act*[6] (the Act) alongside the Complaints and content-based removal notices schemes.

Issues and gaps arising from interplay of the current approach to online safety and the rapid rise and evolution of digital risks in Australia

- A focus on mandatory notice and take down, delivered through the Complaints and content-based removal notices schemes, is:
    - Limited in focus to particular types of content. Not all risky content is covered by the Act, nor could it be. The dynamics of digital risk means that any list of content types subject to notice and take down, no matter how extensive, would become rapidly out of date. Further, not all risks online emerge from content, and notice and take down cannot address these risks.
    - Works 'downstream' after the harm has happened. For this process to be triggered, content has to have been posted online and (largely) seen and already caused harm. It does not require 'upstream' actions that prevent harm in the first instance.
    - Cannot meet the scale of the risk. User generated content systems already generate content prolifically, with estimated for example, of up to 34 million videos posted to TikTok daily.[7] The rise of Generative AI means that more and more complex content

[1]Via the *Enhancing Online Safety for Children 2015* Act https://www.legislation.gov.au/C2015A00024/2017-06-23/text
[2]Via the *News Media and Digital Platforms Mandatory Bargaining Code 2021* https://www.legislation.gov.au/C2021A00021/asmade/text
[3]UK 2023 *Online Safety Act 2023* https://www.legislation.gov.uk/ukpga/2023/50/enacted
[4]EU 2022 *Digital Services Act* https://eur-lex.europa.eu/eli/reg/2022/2065/oj
[5]Canada 2024 *Online Harms Bill 2024* https://www.parl.ca/LegisInfo/en/bill/44-1/c-63
[6]Commonwealth of Australia 2021 *Online Safety Act* https://www.legislation.gov.au/C2021A00076/latest/text
[7]Sarah Anderson 2024 *TikTok Stats and Analytics to Know in 2024* https://www.socialchamp.io/blog/tiktok-stats/

is being easily created.[8] Issuing notices to specific pieces of content cannot meet the sheer scale of these risks.[9]

- The 'systemic' focus in the Act, delivered through the Basic Online Safety Expectations (BOSE) and Online Content Scheme, is:
  - Delivered by voluntary or at best co-regulatory schemes. These do not produce high quality protections for Australians,[10] and can simply be ignored by platforms. They rely on creating 'reputational risks' where platforms violate them but there are limits the 'reputational risk' approach;[11]
  - Limited in focus to a specific set of risks. The desire to stay within the existing mandate of the Act has replicated a narrow focus that fails to address the breadth and scale of online risks Australians now face.

- Transparency powers are limited. The limited focus of the BOSE has 'knock on' consequences; it restricts access to information for the Office of the eSafety Commissioner to request periodic or non period reports only regarding information relevant to the BOSE. This has consequential effects for public transparency processes, which rely on public summaries issued by the Office of the eSafety Commissioner. This does not create the conditions necessary for meaningful public transparency.[12]

- Enforcement powers and fines are vulnerable to dismissal by very large platforms as we have seen play out.[13] Other jurisdictions and other Australian regulators have more significant finding regimes,[14] and in the UK, there are even some potential criminal sanctions associated with online safety reporting.[15]

Figure 1: Inherent issues with Australia's existing online safety framework

[8]For example, an industry blog estimates that more images have been made by Generative AI than taken by photographers in 150 years, speaking to a capacity of generative content to dramatically overshadow current experiences. See Every Pixel Journal 2024 *AI Has Already Created As Many Images As Photographers Have Taken in 150 Years* https://journal.everypixel.com/ai-image-statistics

[9]See for example Europol Innovation Lab 2022 *Facing reality? Law enforcement and the challenge of deepfakes, an observatory report from the Europol Innovation Lab* https://www.europol.europa.eu/cms/sites/default/files/documents/Europol_Innovation_Lab_Facing_Reality_Law_Enforcement_And_The_Challenge_Of_Deepfakes.pdf

[10]Reset.Tech Australia 2022 *How outdated approaches to regulation harm children and young people and why Australia urgently needs to pivot* https://au.reset.tech/uploads/report_-co-regulation-fails-young-people-final-151222.pdf

[11]Tess Bennett 2024 'Social media giants 'no longer fear reputation risks' *AFR* https://www.afr.com/technology/social-media-giants-no-longer-fear-reputation-risks-20240422-p5flls

[12]See for example, Reset.Tech Australia 2024 *Achieving digital platform public transparency in Australia* (forthcoming)

[13]eSafety Commissioner 2023 *eSafety demands answers from Twitter about how it's tackling online hate* https://www.esafety.gov.au/newsroom/media-releases/esafety-demands-answers-from-twitter-about-how-its-tackling-online-hate

[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-andenforcement/fines-andpenalties/)

[15]For more information see UK 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

1. A systemic focus through introducing a duty of care

Many regulatory approaches either hold end-users responsible for harms, either by identifying and making end-users liable for defamation, or by seeking to 'responsiblise' end-users to keep themselves safe, such as by introducing requirements for education or parental approval. While these approaches have merit, they overlook the significant role that digital platforms themselves have in generating and amplifying digital risks.

The most significant factor in generating risk and shaping the risk architecture of the digital ecosystem for Australians is the design and business decisions made by digital platforms. A 'systemic' approach to online safety regulation focuses on this, and ensures that the systems and processes that digital platforms deploy reduce risks and prevent harms. Lorna Woods, for example, describes four key systems and processes that are critical intervention points towards online safety:

- Access to services and content creation
- Discovery and navigation
- User response tools
- Platform response tools[16]

Each of these areas has been shown to create risks for Australian end-users. For example, in previous research, using eating disorder risks as a case study, we have identified how nefarious actors are able to create paid-for content that creates risks, how discovery features such as recommender systems amplify risks and how user-reporting systems do not lead to content take down thereby continuing risks.[17]

Ensuring that digital platforms play their part in reducing the risk architecture requires 'flipping the table' from older models of regulation where end-users shoulder the bulk of the risk and instead placing responsibilities onto digital platforms to keep end-users safe. Learning from international models, placing a duty of care on digital platforms could help to drive the systemic and preventative focus that is urgently needed in Australia.

A duty of care approach is a way to implement systemic regulation that moves the focus beyond the content layer of the digital world to focus on 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 still encouraging users to express themselves.

Focusing on design and operation is important because despite their name, 'platforms' are not entirely neutral, passive transmitters when it comes to content. Intentionally or not, their choice 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.

Duty of care is a familiar model for risk management in Australia, with established models in workplace health and safety. An online, statutory duty of care exists in the UK's *Online Safety Act* ('UK OSA')[18] and is contemplated in draft Canadian legislation, Online Harms Bill.[19]

---

[16]Lorna Woods & Will Perrin 2022 A *modern systems approach to regulating online hate speech* https://cdn.epra.org/attachments/files/4161/original/EPRA_-_Woods_Perrin_hate_speech_systemic_approach.pdf?1656591545
[17]Reset.Tech Australia 2024 *Not just algorithms* https://au.reset.tech/news/report-not-just-algorithms/
[18]UK 2023 *Online Safety Act 2023* https://www.legislation.gov.uk/ukpga/2023/50/enacted
[19]Canada 2024 *Online Harms Bill 2024* https://www.parl.ca/LegisInfo/en/bill/44-1/c-63

We note that the UK's experience in drafting the UK OSA saw proposals for a singular duty of care[20] eventually implemented as a series of overlapping *duties* of care largely regarding illegal content, content that is risky to children and, for larger platforms, content that is risky to adults (see Figure 2). This approach requires distinguishing between different types of content – such as criminal content, content harmful to children and, for larger platforms, content harmful to adults – and then associating specific duties to each type of content.

While inevitably this was the preferred approach for technology companies as it restricts the breadth of obligations, it has created "gaps" in protections for end-users. It is unclear, for example, how the UK OSA is going to address harms arising from overarching abusive designs that do not fall into a particular sort of content, such as dark patterns that deceive users or extended use design techniques deployed at children, for example.

It also introduces an unusual paradox that stops the obligations being truly systemic (and preventative). A singular duty of care approach acknowledges that systems are developed and business decisions are made *before* platforms are populated with content. Platforms decide how their content recommender systems will work, or how their moderation teams will be staffed etc., without knowing what content they will recommend or moderate each day. Complying with a singular duty of care obligation means that platforms are encouraged to safeguard their systems before harm happens. Implementing duties of care tied to particular sorts of content, in contrast, requires platforms to risk assess their systems after they are 'populated' with designated content, or after harm has happened. This seems at odds with the sort of 'upstream' and systemic approach that a duty of care seeks to enable.

Implementing duties of care rather than a singular duty of care moves the regulation away from a focus on the systems and back into specifying particular types of content. This skews the focus of compliance towards a content-first rather than a systems-first approach. This was present in much of the Parliamentary debate in the UK, which became very focused on what content would be removed and what would not, and we can see this tension emerging in the Australian political dialogue regarding content-focused digital regulation.[21] This is not necessary nor desirable. A systemic focus would enhance rather than erode public trust in the Act.

Further, maintaining a focus on systems, through a singular duty of care enhances expression. By placing obligations on digital platforms to safeguard (and as we argue below, be transparent about) the inner workings of their systems that shape public discourse, both safety and expression are enhanced. Limiting obligations instead to specific duties tied to particular types of content undoes this.

A duty of care could apply to all digital platforms with Australian end-users.

[20]Lorna Woods 2019 *Online harm reduction – a statutory duty of care and regulator* https://ssrn.com/abstract=4003986

[21]See for example Sky News 2023 'ACMA agency being given position as the 'official censor of the internet' *Sky News* https://www.skynews.com.au/opinion/chris-kenny/acma-agency-being-given-position-as-the-official-censor-of-the-internet/video/ac27a65a775b137318dd0954851312a6

Pluralised duties in the UK Online Safety Act

All user-to-user systems have duties regarding:
- Illegal content risk assessments;
- Illegal content;
- Content reporting;
- Complaints procedures;
- Freedom of expression and privacy, and;
- Record keeping and review.

All services likely to be accessed by children have duties regarding:
- Children's risk assessments, and;
- Protecting children's online safety.

The largest online services also have additional duties regarding:
- Adult risk assessment duties;
- Duties to protect adults' online safety;
- Duties to protect content of democratic importance, and;
- Duties to protect journalistic content.

Figure 2: Duties of care in the UK Online Safety Act[22]

2. A comprehensive focus through mandating risk assessment requirements

Once responsibility has been placed onto digital platforms to safeguard end-users, requirements to produce risk assessments could introduce a comprehensive focus into the regulatory framework. This approach has strong international precedent; requirements to produce risk assessments for systemic risk on digital platforms exist in both the EU's *Digital Services Act* (DSA)[23] and the UK's OSA.

Currently risk assessments are part of the Australian BOSE, although they are suggested as an example of a reasonable step to address specific risks covered by the BOSE. They are neither mandatory nor comprehensive. In addition, the Office of the eSafety Commissioner has created a world-leading Safety By Design assessment tool, which forms as guidance and advice for digital product developers.[24] This product has significant strengths, but it is a self-assessment tool linked to a set of safety risks, and was not designed to support regulatory enforcement.

Requirements to produce risk assessments could ensure that platforms must adequately review and identify the risks that their systems and processes create. As the Centre on Regulation in Europe describes, risk assessment activities begin with a comprehensive mapping activity that identifies the ecosystem that platforms operate in, the roles and behaviours of users, business decisions made by platforms and how these interface to produce risks.[25] That is, risk assessments have the capacity to encourage digital platforms to think comprehensively about how their platforms can create or amplify risks.

Failures to adequately identify risks at the risk assessment stage can lead to significant consequences. We have already seen under the DSA that the European Commission has commenced enforcement actions against platforms that have failed to adequately identify risks

[22]UK 2023 *Online Safety Act 2023* https://www.legislation.gov.uk/ukpga/2023/50/enacted
[23]EU 2022 *Digital Services Act* https://eur-lex.europa.eu/eli/reg/2022/2065/oj
[24]Office of the eSafety Commission 2023 *Assessment tools* https://www.esafety.gov.au/industry/safety-by-design/assessment-tools
[25]Sally Broughton & Micova Andrea Calef 2022 *Elements For Effective Systemic Risk Assessment Under The DSA* https://cerre.eu/wp-content/uploads/2023/07/CERRE-DSA-Systemic-Risk-Report.pdf

in the first instance. For example, the European Commission has opened formal proceedings against TikTok for failing to adequately identify risks from their system, including the stimulation of behavioural addiction and harms from "rabbit-hole" effects for minors.[26]

In the absence of a pre-existing, EU-wide 'duty of care' principle in European law, requirements for risk assessments in the DSA are used to identify and shape the nature of the obligations on digital platforms. Specifically, the DSA requires platforms to assess against risks of:

- The dissemination of illegal content (where illegality is defined by the laws of member state countries);
- 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, and;
- Negative effects on gender-based violence, public health, children's wellbeing and serious negative consequences to people's physical and mental wellbeing.[27]

The DSA specifically describes how these risks should be considered in systemically focussed risk assessments, but noting that platforms must consider the risks posed by (but not limited to) the following systems:

- Recommender systems and other algorithms;
- Content moderation systems;
- Terms and conditions and their enforcement;
- Systems for selection and presenting advertisements, and;
- Data related practices of the provider.[28]

Risk assessments in the EU & UK must be sent to regulators on a regular basis, for regulators to assess adequacy and compliance. In the EU, summaries of these risk assessments are expected to be made public, to additionally enhance public transparency.

If there is a desire within a revised Australian framework to clarify core systemic focusses for technology companies, there is the capacity to do this in shaping minimum requirements for risk assessments. (This would be preferable to reducing the scope of accountability from a duty of care to specific duties of care, for reasons described above). Building on existing Australian requirements, and harmonising with EU requirements to reduce burden on platforms, Australian minimum requirements for risk assessments could include:

<u>Risk assessments must consider at least the following risks:</u>

- The dissemination of illegal and harmful materials, as already defined in the *Online Safety Act*;
- The dissemination of online scams;
- Negative effects on electoral processes and public security;
- Negative effects to civil and political rights, such as political freedoms, freedom of opinion and expression, and;
- Negative effects on gender-based violence, children's best interests, public health and serious negative consequences to people's physical and mental wellbeing.

<u>Risk assessments must consider at least the following systems:</u>

- Recommender systems and other algorithms;
- Content moderation systems;
- Terms and conditions and their enforcement;
- Systems for selection and presenting advertisements, and;

[26]European Commission 2024 *Commission opens formal proceedings against TikTok under the Digital Services Act* https://ec.europa.eu/commission/presscorner/detail/en/ip_24_926
[27]Article 34, EU 2022 *Digital Services Act* https://eur-lex.europa.eu/eli/reg/2022/2065/oj
[28]Article 34, EU 2022 *Digital Services Act* https://eur-lex.europa.eu/eli/reg/2022/2065/oj

- Data related practices of the provider where they create safety risks.

Risk assessment requirements could be 'tiered' and applied only to platforms that have significant numbers of Australian end-users.

3. A preventative focus through mandating risk mitigation requirements

The responsibility to identify a comprehensive, systemic set of risks can become preventative where digital platforms are required to actively mitigate and minimise the likelihood and severity of these risks. This way, platforms can be incentivised to create changes that prevent harms occuring the first instance. In this sense, as the idiom goes, risk mitigation measures are the equivalent of 'placing a fence at the top of a cliff rather than ambulances at the bottom'.

Again, there is strong international precedent for risk mitigation requirements. The EU's *Digital Services Act*[29] and the UK's OSA places obligations on platforms to mitigate identified risks, and Canada's Online Harms Bill also places obligations on platforms to mitigate risks aligning with their duties. Currently risk assessments which include risk mitigation measures are part of the Australian BOSE, although they are suggested as an example of a reasonable step in response to a range of risks covered by the BOSE and are not mandatory.

We have seen requirements for risk mitigation measures begin to take effect overseas. For example, the European Commission has opened formal proceedings against Meta for failing to adequately identify risk mitigation measures to curb harms to minors[30] and for failing to adequately adopt mitigation measures against visibility around political content and illegal content flagging, among others.[31]

The DSA specifically outlines a set of 'mitigation measures' that could be expected from digital platforms, such as:

- Changing the design, features or functioning of their services, including their online interfaces;
- Changing terms and conditions and their enforcement;
- Changing content moderation processes;
- Testing and changing algorithmic systems, including recommender systems;
- Changing advertising systems, including the way ads are targeted at or presented to people;
- Improving internal business processes to maximise safety;
- Collaborating with other digital services;
- Taking targeted measures to improve child safety, such as age assurance or parental control tools, and;
- Ensuring evidence about potential illegal activities is stored and reported in helpful ways to law enforcement.[32]

Australian expectations could harmonise with EU requirements to reduce compliance burden on platforms. This would introduce a strong mechanism that encourages platforms to implement preventative measures and allows regulators to meaningfully interrogate proposed measures while they are still risks rather than actualised harms.

Regulators should also be empowered to draft industry standards about what each of these risk assessments should look like, and what adequate risk mitigation measures should be.

---

[29]EU 2022 *Digital Services Act* https://eur-lex.europa.eu/eli/reg/2022/2065/oj
[30]European Commission 2024 *Commission opens formal proceedings against Meta under the Digital Services Act related to the protection of minors on Facebook and Instagram* https://ec.europa.eu/commission/presscorner/detail/en/ip_24_2664
[31]European Commission 2024 *Commission opens formal proceedings against Facebook and Instagram under the Digital Services Act* https://ec.europa.eu/commission/presscorner/detail/en/IP_24_2373
[32]Article 35, EU 2022 *Digital Services Act* https://eur-lex.europa.eu/eli/reg/2022/2065/oj

4. Enhancing trust and safety through transparency

Regulating for transparency helps to address the power asymmetry of large digital platforms by rendering visible some of the information that the public and regulators need to understand online risks. This enables both individuals and regulators to respond, from allowing consumers to make informed choices about the use of platforms to allowing regulators to take action. Transparency is not a silver bullet, but alongside other systemic, comprehensive and preventative measures, it can help to redress the harms of digital platforms on individuals and society.

Current Australian measures for transparency in the online safety framework emerge from requirements in the BOSE. Under the BOSE, the Office of the eSafety Commissioner has the powers to request a range of information from platforms via 'transparency notices'.[33] These powers include:

- Requiring platforms to provide periodic reports, ranging from 6 monthly to 24 months about compliance with the BOSE. The reports would need to provide information about the extent to which the platform complied with the BOSE in general or specific elements of the BOSE,[34] in a manner and form that is specified by the Commissioner.[35] To date, we are unaware of any platform that has been required to produce periodic reports.

- Requiring non-periodic reporting about compliance, where each platform provides a particular type of service (like search engines, or online messaging services) about compliance with one or more elements of the BOSE, in a manner and form that is specified by the Commissioner.[36] To date, there have been three 'rounds' of non-periodic requests made to platforms.[37]

While responses to these notices are sent directly to the Office of the eSafety Commissioner, the Commissioner is empowered to publish a statement regarding the platforms' periodic and non-periodic reports on their website that delivers a subsequent public transparency function.[38] Platforms have not always adequately responded to these requests,[39] and these are modest transparency requirements compared with overseas regulatory benchmarks.

Internationally, transparency requirements are stronger in other markets. For example the DSA introduces five key types of public transparency measures:

- Annual risk assessments which are released in a summary form to the public after a period of time;
- Annual transparency reports which are highly prescriptive and share detailed data about the functioning of platforms;
- Annual, independent audits that provide independent oversight of platform drafted reports;
- Data portals, including ad repositories and content moderation data, and;

[33]eSafety Commisioner 2024 *Responses to transparency notices*
https://www.esafety.gov.au/industry/basic-online-safety-expectations/responses-to-transparency-notices
[34]Commonwealth of Australia *Online Safety Act* 2021 Division 3(A) 49 3
https://www.legislation.gov.au/C2021A00076/latest/text
[35]Commonwealth of Australia *Online Safety Act* 2021 Division 3(A) 49 2
https://www.legislation.gov.au/C2021A00076/latest/text
[36]Commonwealth of Australia *Online Safety Act* 2021 Division 3(A) 56 2
https://www.legislation.gov.au/C2021A00076/latest/text
[37]Available at eSafety Commissioner 2024 *Responses to transparency notices*
https://www.esafety.gov.au/industry/basic-online-safety-expectations/responses-to-transparency-notices
[38]Commonwealth of Australia *Online Safety Act* 2021 Division 3(A) 59 2
https://www.legislation.gov.au/C2021A00076/latest/text
[39]See for example, See *X Corp v eSafety Commissioner* (VID956/2023), status available at:
https://www.comcourts.gov.au/file/Federal/P/VID956/2023/actions

- Researcher access to public interest data.[40]

Likewise, the UK OSA introduces two key public transparency measures:
- Annual risk assessments which will be published in summary for the public, and;
- Annual transparency reports.[41]

There is even policy discussion around introducing transparency requirements in the US, via independent research in the proposed Kids Online Safety Act.[42]

Building on these examples, we have identified five public transparency measures that we think could be adopted within the online safety framework in Australia:

1. Risk assessments, which are detailed documents sent directly to regulators, but also made available in summary version to the public after a reasonable period of time;
2. Annual transparency reports, which are detailed and prescriptive (see Figure 3 for examples of potential prescriptions, which also highlights how this measure would enhance trust);
3. Annual independent audits. An independent expert or 'skilled person' should be required to review both the platforms' risk assessments and transparency notices;
4. Data portals. This would include searchable ad repositories, and data about content moderation decisions. Data from the EU suggests that most content moderation decisions are made on the basis of violations of terms of service, such as self harm policies or dangerous challenge policies, not about issues related to civic discourse or elections.[43] This suggests inclusion in the online safety framework would be appropriate, and;
5. Researcher access requirements. Vetted researchers in Australia should be able to request data. Like in the EU, requirements for an Australian vetted researcher could include:

    - Affiliation to a research organisation, including academic and third sector research organisations;
    - Researchers or at least the lead researcher should be an Australian resident, and;
    - Non-commercial purpose limitations.

    Suitable research projects should be provided with data. A suitable project proposal would including information demonstrating that:

    - The research fits the ambitions of the *Online Safety Act* and how it is broadly of public benefit. This does not include data about trade secrets;
    - Funding for the research is fully disclosed;
    - Access to the specific data requested, and the indicated timeline indicated, is necessary and proportionate to the purposes of the research;
    - Data security and confidentiality requirements, as well as personal data safety requirements, will be fulfilled, and;
    - The research results will be made publicly available free of charge within a reasonable period after the completion of the research. The process for requesting data could be managed by the Australian Communications and Media Authority, the Office of the eSafety Commissioner or other appointed independent organisation. In addition, existing data and data tools like APIs should be made available to Australian researchers for free.

[40]See for example, Reset.Tech Australia 2024 *Achieving digital platform public transparency in Australia* (forthcoming)
[41]See for example, Reset.Tech Australia 2024 *Achieving digital platform public transparency in Australia* (forthcoming)
[42]US Senate Committee on Commerce, Science, and Transportation 2023 *S.1409 - Kids Online Safety Act* https://www.congress.gov/bill/118th-congress/senate-bill/1409/text
[43]European Commission 2024 *DSA Transparency Database* https://transparency.dsa.ec.europa.eu/dashboard

Public transparency requirements could be 'tiered' and applied only to platforms that have significant numbers of Australian end-users.

These five public transparency measures need to exist alongside information gathering powers afforded to the Office of the eSafety Commissioner, such as expanding powers around transparency notice requests. These are detailed in our response to Question 17.

Potential prescriptions for online safety transparency reports

- Metrics on the design, features, or functioning of services, e.g.:
  - Data around internal safety tests made of features and systems conducted, including a description of tests and outcomes, and nature of adaptations made as a result that affect Australian end-users;
  - Changes to community guidelines and terms of service for Australian end-users, and;
  - Human resources dedicated to trust and safety, including information about; numbers located within Australia; numbers dedicated to Australian safety issues; qualifications and training; and support.

- Problematic use metrics, e.g.:
  - Number of adult users demonstrating problematic over-use, and data about average and median use times;
    - Number of push notifications sent to these users on average per day (in app and out of app);
  - Number of child users (under 18 years) demonstrating problematic over-use, and data about average and median use times;
    - Number of push notifications sent to these users on average per day (in app and out of app);
  - Number of child users (under 18 years) accessing the platforms between 10pm and 6am in their time zone, and data about average overnight usage, and;
  - Estimates of number of users under the minimum age of use according to the terms of service, and data about average detection and response to these accounts.

- Child sexual exploitation and abuse metrics:
  - Numbers of adult users blocked for contact with minors, and data about response times and previous reportings of users;
  - Numbers of adult users reported by minors, and data about responses and response times, and;
  - Number of CSAM reports made, and data about response and response times.

- Online scam metrics:
  - Number of scam posts reported on the platform, including data about detection methods (organic or user-report), average engagement and responses including average response time.

- Content moderation metrics relating to online safety, including impact on Australian businesses and pages. e.g.:
  - Number of organic content measures (i.e. how much content platforms proactively detected) that violated their community guidelines; by violation type (e.g. violated self harm policy, violated dangerous challenges policy, pornography etc.); amount detected by automated means; amount detected by human moderators; median, average and max time to detect these, and final response; business specific metrics, e.g.:
    - Number of organic business entity measures (i.e. how many Australian business accounts were removed or restricted as a result of organic content moderation);

        - Number of organic entity measures (i.e. how many Australian pages or products were removed or restricted as a result of organic content moderation);
    - Number of user-reported content measures (i.e. how much content was reported to the platform by Australian end-users) that violated their community guidelines; by violation type; median, average and max time to detect these; response; number of challenges against response; final outcomes; business specific metrics, e.g.:
        - Number of business entity measures following user-reporting (i.e. how many Australian business accounts were removed and restricted after user-reporting)
        - Number of entity measures following user reporting (i.e. how many Australian pages or products were removed and restricted after user-reporting);
    - Number of 'trusted-flagger' content measures (i.e. who are platforms trusted flaggers in Australia, like eating-disorder experts, suicide prevention experts or Australian fact-checkers); how much content was acted on by a platform as a result of trusted flaggers; amount reported to platform; by violation type; amount of 'identical or near indenitical' content subsequently detected by automated means; median, average and max time 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, and;
    - 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 are they required to review); language addressed.

- Measures against misuse such as number of Australian end-users' accounts suspended or deleted and why; number of challenges, and final outcome.

- Number of Australian end-users monthly, including breakdowns by under 18 and over 18 years.

Figure 3: Potential prescriptions for online safety transparency reports

5. Enhancing enforcement to ensure meaningful change

A duty of care model, risk assessments, risk mitigation and transparency measures will not lead to demonstrable improvements in online safety unless these regulations are enforced. As we have seen with the current co-regulatory and voluntary approach, where platforms have an outsized role in setting their own standards, safety is not maximised.

International regulators have a range of enforcement powers that are not currently available to the Office of the eSafety Commissioner, to compel redress. These regulators can ensure platforms change and improve safety standards. To be clear, regulators must be able to outline what they believe appropriate risk assessment and risk mitigation measures should be, and need to be able to take enforcement actions where platforms fail to make required improvements. Available enforcement actions should include:

- The ability to issue significant fines for failures to meet required improvements. Figure 4 highlights the scale of the fining regime available to comparable regulators.

- Strong, last resort measures designed to prevent platforms from ignoring regulators' requests. For example:

    - Under the DSA where the failures are significant and persistent, and attempts at engagement have failed, regulators can 'turn off' services. Specifically, the DSA outlines

that if an "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" regulators can work with domestic courts to order temporary restrictions of access.[44]

- Alternatively under the UK OSA, with the agreement of the courts, Ofcom can require payment providers, advertisers and internet service providers to stop working with a site, preventing it from generating money or being accessed from the UK.[45]
- In extreme cases in the UK, criminal sanctions can be issued to senior management if transparency measures are not met. The UK OSA requires companies to identify senior managers who are liable for responding to information notices. Failure to comply with an information notice request is a criminal offence.[46] These measures stand in stark contrast to Australian enforcement powers, where requests for information have been ignored and fines of $610,500 issued.[47]

Potential fining regimes available for the online safety framework

- Under the UK OSA, companies can be fined up to £18 million or 10% of their qualifying worldwide revenue, whichever is greater.
- Under the DSA, companies can be issued penalties of up to 6% of global annual turnover for failure to effectively mitigate risks, of up to 1% of global annual turnover for supplying incomplete or misleading information as part of meeting transparency obligations.
- In Australia, regulators in the adjacent domains of consumer protection and financial services have comparable fining abilities. For example the ACCC can fine up to 10% of annual turnover for franchising violations[48] and ASIC can fine up to 10% of annual turnover, capped at $782.5 million, for violations of ASIC administered legislation.[49]

Figure 4: Fining regimes available to other regulators

[44]EU 2022 *Digital Services Act* https://eur-lex.europa.eu/eli/reg/2022/2065/oj, Article 51(3)
[45]UK Department for Science, Innovation & Technology 2024 *Online Safety Act: explainer* https://www.gov.uk/government/publications/online-safety-act-explainer/
[46]For more information see UK 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
[47]Georgie Hewson 2023 'Australia's eSafety commission fines Elon Musk's X $610,500 for failing to meet anti-childabuse standards' *ABC* https://www.abc.net.au/news/202310-16/social-media-x-finedover-gaps-in-child-abuseprevention/102980590
[48]ACCC nd *Fines and penalties* https://www.accc.gov.au/business/compliance-and-enforcement/fines-and-penalties
[49]ASIC 2023 *Fines and Penalties* https://asic.gov.au/about-asic/asic-investigations-and-enforcement/fines-and-penalties/

2. Response to the *Issues Paper*

We warmly welcome the *Statutory Review of the Online Safety Act: Issues Paper* and the comprehensive direction of the review. The commitment to ensuring an effective, future-proofed regulatory framework for online safety and the breadth of the review are especially welcome. Enhancing platform accountability including through a duty of care; the introduction of the children's best interests requirement; the opportunity to draw on international models of best practice; and the intent to investigate effective enforcement are urgent and necessary areas of focus. This review is timely and necessary.

We note that the issues paper outlines a strong desire to move beyond a content focus and notes a range of risks that also emerge from conduct and contact. For example, Part 6 cites the 'Three C's' (3C) typology of harm when describing the types of harms Australians experience; content, conduct and contact. Likewise, the World Economic Forum's *Typology of Online Harms* is frequently cited across the document, which also uses the 3Cs typology, but groups these into categories of risks (e.g. the content, conduct and contact risks associated with threats to personal safety, or threats to privacy).[50] The move beyond content is welcome, but neither typology of harms is wholly compatible with a systemic focus. 3C harms are largely formulated to describe how bad actors use technology to create risks (such as bad actors who doxx women creating a content based threat to privacy, or bad actors who use digital technology to groom children creating a contact risk to personal safety). Partly for this reason, the 3Cs framework was expanded in 2021 to include a 4th C (contract risks) and to render visible the extent of cross-cutting risks.[51] Cross-cutting risks and contract risks create the space to explore the role of the digital architecture and platforms themselves in creating risks, while still maintaining a focus on 'bad actors' and individual users. While this is helpful to an extent, both 4C and 3C typologies foreground risks that are not systemic in nature, so we recommend using them with caution. The most effective digital regulations have emerged where the focus has remained on the systemic risks that platforms create, and therefore platforms can straightforwardly mitigate against.

Our concern is that without a systemic focus, a revised *Online Safety Act* runs the risk of becoming a 'Christmas tree bill', where long lists of content types and conduct risks are designated in codes or regulations. 'Decorating' the Act with endless lists of 'bad things' and 'bad behaviours' will both ensure that the Act is effectively out-of-date the moment it is published (because digital risks are inherently dynamic and move at pace), and; will create a list of disjointed obligations with all sorts of contested definitional issues that will need resolving. A more effective approach builds on international experiences to create an overarching obligation towards user safety, and requires platforms to regularly risk assess and risk mitigate to ensure these obligations are met.

Overall, we believe that a range of proposals explored in the *Issues Paper* have the potential to create a radically improved online safety framework across Australia. Below, we address the specific questions from the *Issues Paper*.

[50]World Economic Forum 2023 *Toolkit for Digital Safety Design Interventions and Innovations: Typology of Online Harms* https://www3.weforum.org/docs/WEF_Typology_of_Online_Harms_2023.pdf
[51]Sonia Livingstone & Mariya Stoilova 2021 *The 4Cs: Classifying Online Risk to Children* https://doi.org/10.21241/ssoar.71817

Part 2: Questions 1-7

Q2. Sections of online sector regulated:

The breakdown of industry into eight subsections is a complicated necessity of a detailed 'notice and take down' approach. Systemic regulation somewhat avoids the need to get stuck in complicated industry definitions by instead requiring bespoke risk assessments and risk mitigations to demonstrate compliance with a duty of care for each specific platform. This approach is desirable for two key reasons:

- Firstly, growing vertical integration and monopolistic practices make classifying platforms a complicated task. For example, it is unclear to us if Whatsapp is best considered as a 'relevant electronic service' in Australia rather than a 'social media service'. Classification based on functionality is complex; while Australians largely use Whatsapp as a messaging service, the platform provides a Status feature (which is functionally the same as Instagram Stories) and allows members to join Communities or Groups (functionally the same as Facebook groups). It is further unclear what would happen if Australians began to use these functionalities at the same rate as users do in other parts of the world. Would Whatsapp then change classifications and what would be the tipping point?

- Secondly, because of the blurred lines of functionality, digital platforms are to a limited extent able to go 'code shopping'. Companies must classify themselves according to their predominant purpose, but this is subjective and the classification is ultimately left up to the platform.[52]

Broad coverage of all digital platforms, with specific obligations for larger online platforms, such as those with Monthly Average Users (MAU) representing 10% of Australian population, or 10% of Australia's child population, would be more straightforward.

Q4. Strengthened and enforceable BOSE:

Safety expectations should be enforceable as they are in other jurisdictions, such as the EU and UK.

Without enforceability, safety expectations become in effect voluntary. For example, the BOSE clearly states that the default privacy and safety settings for children must be robust and set to the most restrictive level,[53] and that a child is someone under 18 years of age.[54] However, this is not what happens in Australia. Because the BOSE are not enforceable, and because we rely on industry-drafted codemaking processes, the actual safety standard in Australia is default privacy settings and safety standards to the most robust and restrictive only for those under the age of 16 years (see Figure 5).

---

[52]See also concerns noted in Alannah & Madeline Foundation 2022 *Online Safety Codes: A submission by the Alannah & Madeline Foundation*
https://www.alannahandmadeline.org.au/uploads/main/Online-Safety-Codes.pdf
[53]Minister for Communications 2024 *Online Safety (Basic Online Safety Expectations) Amendment Determination 2024*
https://www.infrastructure.gov.au/sites/default/files/documents/online-safety-bose-amendment-determination-2024.pdf, Schedule 1 Amendments, 3, which states that 'if a service or a component of a service (such as an online app or game) is likely to be accessed by children (the *children's service*) – ensuring that the default privacy and safety settings of the children's service are robust and set to the most restrictive level'
[54]Commonwealth of Australia 2021 *Online Safety Act 2021*
https://www.legislation.gov.au/C2021A00076/latest/text, which defines child as 'an individual who has not reached 18 years'

How safety settings and default privacy settings are handled in the Social Media Services Online Safety Code (Class 1A and Class 1B Material)

'Definitions:
- Australian child: Australian child means an Australian end-user under the age of 18 years.
- Young Australian child: young Australian child means an Australian end-user under the age of 16 years.'

Minimum compliance measures for Tier 1 social media services:

'A provider of a Tier 1 social media service that permits a *young* Australian child to hold an account on the service must at a minimum:

a) have default settings that are designed to prevent a *young* Australian child from unwanted contact from unknown end-users, including settings which prevent the location of the child being shared with other accounts by default; and
b) easy to use tools and functionality that can help safeguard the safety of a *young* Australian child using the service.'

Figure 5: Excerpts from the Social Media Services Online Safety Code (Class 1A and Class 1B Material) codes, (emphasis on young added)[55]

Q5. Code drafting processes:

The process of allowing industry to draft safety codes has demonstrably failed to improve safety standards for Australians. As we have highlighted previously:[56]

- Co-regulation does not meet community expectations and the public overwhelmingly wants regulation drafted by regulators or legislators. Working with YouGov, Reset.Tech Australia commissioned a Dec 2022 poll of 1,508 Australian adults and found that only 21% trusted the social media industry to write their own codes, and the majority said they would prefer if independent regulators drafted any safety and privacy codes. Likewise, an April 2022 poll of 506 Australian teenagers by YouGov found that only 14% said they trusted social media companies to 'write the rules' when it comes to privacy.

- Co-regulation demonstrably leads to weaker protections. Exploring the online safety codes written by industry for Australia to similar codes written by regulators elsewhere in the world, it becomes apparent that co-regulation offers weaker protection. Three examples are documented in our previous work:
    - Young people's accounts must be set to "maximum privacy" up until the age of 18 years according to regulator drafted codes in the UK and Ireland, but only up until the age of 16 years under Australian codes. This leaves Australian 16 and 17 year olds less protected;
    - Children's precise location data can be collected in Australia, creating real safety and privacy risks. Regulator-drafted codes in the UK and Ireland prevent the collection of unnecessary children's location data, and;
    - Child sexual abuse reporting requirements are less clear and rigorous in industry-drafted codes than legislator drafted protections in the UK.

[55]Communications Alliance & Digi 2023 *Schedule 1 – Social Media Services Online Safety Code (Class 1A and Class 1B Material)*
https://onlinesafety.org.au/wp-content/uploads/2023/06/230616_1_SMS-Schedule_REGISTERED-160623.pdf
[56]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/

- Co-regulation is inappropriate given the level of risk technology creates, and the behaviour of dominant tech companies. Technology creates significant risks for the Australian community, including public health risks, and there is a track record of undermining emerging regulations among the tech sector.[57]

The weakness of the current online safety codes should not be understood as a one-off peculiarity. The *Australian Code of Practice on Disinformation and Misinformation*, authored by industry, has also required subsequent intervention from the ACMA to strengthen,[58] and in the EU too voluntary codes ultimately had to be subsumed within their *Digital Services Act* because they failed to deliver change.[59] Where industry authors codes, weaker protections are offered and regulators inevitably have to step up. The issue is that Australians continue to be harmed during the delay while we wait for co-regulation to fail.

We do not believe that industry has the sufficient expertise nor the right incentives to prioritise end-users' safety, and this should instead be left to regulator drafted processes. We note that this could be in keeping with revisions to the *Privacy Act*, where proposals have been made to allow the OAIC to draft codes where it is in the public interest to do so, and 'where there is unlikely to be an appropriate industry representative to develop the code.'[60] Industry representative, saddled with significant conflicts of interest and track records of routine undermining of safety considerations, are far from appropriate authors of codes.

Q6. Service providers terms of service:

Platforms' terms of service are important but are voluntary and currently poorly enforced. Because they are voluntary, there is little incentive for platforms to improve terms of service (i.e. offer stronger safety and privacy protections for end-users), without regulatory demand. Globally, we see terms of service improve where regulations require. For example, we have seen:

- Major platforms improve their terms of service to improve children's safety in the UK following the introduction of the *Age Appropriate Design Code*.[61]

- Very large online platforms change their terms of service for EU users in anticipation of the DSA, such as changing terms around content moderation disputes or recommender systems engagement.[62]

Further, without oversight or enforcement mechanisms, we also see platforms fail to enforce their current terms of service regarding user safety. For example, despite having strong terms of service regarding pro-eating disorder content, we see major platforms fail to deliver on this, for example failing to remove it when it is reported and recommending it to children in their feeds.[63]

Leaving online safety simply to each platform's terms of service will fail to lead to demonstrable safety improvements for Australians. Mandatory regulation needs to shape the basic safety standards that Australians should enjoy through platforms' guidelines and terms.

---

[57]Rys Farthing and Dhakshayini Sooriyakumaran 2021 'Why the Era of Big Tech Self-Regulation Must End' *AQ Magazine* https://www.jstor.org/stable/27060078
[58]Office of the Hon Michelle Rowland MP, Minister for Communications 2023 *New ACMA powers to combat harmful online misinformation and disinformation*
https://minister.infrastructure.gov.au/rowland/media-release/new-acma-powers-combat-harmful-online-misinformation-and-disinformation
[59]For example, obligations under the self-regulatory *Code of Practice on Disinformation* (2018) were found to be inadequate, and replaced by obligations within the *Digital Services Act*
[60]Office of the Attorney General 2023 *Government Response: Privacy Act Review Report*
https://www.ag.gov.au/sites/default/files/2023-09/government-response-privacy-act-review-report.PDF
[61]Steve Wood 2024 *Impact of regulation on children's digital* lives
https://eprints.lse.ac.uk/123522/1/Impact_of_regulation_on_children_DFC_Research_report_May_2024.pdf
[62]European Commission 2024 *The impact of the Digital Services Act on digital platforms*
https://digital-strategy.ec.europa.eu/en/policies/dsa-impact-platforms
[63]Reset.Tech Australia 2024 *Not just algorithms* https://au.reset.tech/news/report-not-just-algorithms/

Q7. Obligations based on risk and reach:

Defining obligations on sector definitions, or content definitions, is a complex process that leads to gaps. Instead, a focus on 'tiering' obligations based on a platform's risk appears to deliver a more flexible approach that balances safety with regulatory burden.

For comparison:

- In the EU, the DSA introduces a category of Very Large Online Platforms (VLOPs) and Very Large Online Search Engines (VLOEs), defined as those with 45 million MAUs,[64] which was roughly 10% of the EU member state population when the DSA was passed.

- In the UK, the UK OSA creates three different categories of service regulation, with categorisation being determined by a combination of functionality and number of users. While these categories are yet to be agreed, Ofcom's advice to Government was for category 1 service to be defined as having either 34 million users (50% of the population) and uses a content recommender system, or has 7 million users (10% of the population), uses a content recommender system and allows users to repost content. Recommendations around category 2 definitions repeat the 10% threshold, while category 3 introduces a 5% threshold.[65]

A streamlined approach that bypasses content for a singular duty of care most likely negates the need to use a combination of functionality and user base to 'tier' obligations. A definition of a large platform as a platform that has 10% of Australia's population as MAUs feels like an appropriate level to designate additional requirements.

However, we would also recommend an additional threshold for services that children are likely to access. Where a service is likely to be accessed by children, an additional threshold of 10% of Australia's 13-17 year old population, or under 18 years population should be applied (depending on the accessibility of the service). Otherwise, a service could have the entirety of Australia's secondary school population using it, but because of the demographic shape of the population, this service may not reach the threshold.

Part 3: Questions 8-16

Reset.Tech does not have research based evidence about the efficacy of the complaints system, so we defer to those who have been using or researching the system.

Based on our understanding of the digital risk architecture in Australia however, we can note that the public facing complaints mechanism does not allow complaints about breaches of the BOSE, nor for 'super complainants' to address broader issues. In comparison, the UK OSA allows super complaints to be made to the regulator, allowing everyone from children's groups to free-speech groups to bring evidence to the regulator.[66] Likewise, the European Commission

[64]European Commission 2024 *Very large online platforms and search engines* https://digital-strategy.ec.europa.eu/en/policies/dsa-vlops
[65]Ofcom 2024 *Categorisation: Advice submitted to the Secretary of State* https://www.ofcom.org.uk/__data/assets/pdf_file/0023/281354/Categorisation-research-and-advice.pdf
[66]UK Department for Science and Innovation 2023 *Children's charities and free speech groups could be allowed to submit super-complaints to Ofcom to keep internet safe* https://www.gov.uk/government/news/childrens-charities-and-free-speech-groups-could-be-allowed-to-submit-super-complaints-to-ofcom-to-keep-internet-safe and UK Department for Science and Innovation 2023 *Super-complaints: eligible entity criteria and procedural requirements* https://www.gov.uk/government/consultations/super-complaints-eligible-entity-criteria-and-procedural-requirements

works in collaboration with civil society organisations to gather evidence about systemic risks on platforms.

Reset.Tech does not have research based evidence around age assurance for restrictive access services, but based on our work around children's privacy, but we can note that:

- Requirements to assure age should be privacy preserving and respect children's right to access age appropriate digital services. That is, it would be a detrimental outcome if young people were blocked from the digital world in a broad way, rather than being blocked in a targeted way from accessing restricted access services.

- Age assurance processes can and should vary depending on their intent. For example, the age assurance mechanisms that are appropriate to ensure that children's accounts have safety settings 'turned on', or their data is better protected, are not the same mechanisms that will be needed to prevent children accessing pornography. The first aims to minimise false negatives (i.e keep children *in* safer services), and the latter need to minimise false positives (i.e. keep children *out* of a service). Processes for age assurance processes should be determined to *either* minimise false positives or false negatives depending on what is in children's best interests.

Part 4: Questions 17-20

Q17: Investigation, information gathering and enforcement powers

Alongside public transparency measures, regulators need to have the power to request relevant information in ways that cannot be ignored or overlooked. Comparable powers overseas include:

- Powers to request information, take interviews and statements, conduct inspections and audits (i.e. enter premises and review records), arising in both the DSA[67] and UK OSA;[68]
- The ability to request ad-hoc expert review, or a skilled-person's report arising from the UK OSA,[69] and;
- Anticipated: the power to compel the disclosure of information specifically about the use of service by deceased child users in the UK.[70]

In the Australian context, the ability of the Office of the eSafety Commissioner to request periodic and non-periodic reports regarding compliance with the (revised) *Online Safety Act* could be meaningfully enhanced by powers to take interviews and statements, conduct inspections and request ad-hoc expert review.

Q18: Australian penalties regime

As figure 4 highlights, the fining regime available to the Office of the eSafety Commissioner is comparably inadequate. Enhanced civil penalty powers, set to 10% of global annual turnover would be more commensurate with the scale of the industry and the severity of the risks this industry poses to Australians.

Q19: Enforcement against overseas service providers

[67]EU 2022 *Digital Services Act* https://eur-lex.europa.eu/eli/reg/2022/2065/oj, article 67-69
[68]UK 2023 *Online Safety Act 2023* https://www.legislation.gov.uk/ukpga/2023/50/enacted, sections 105-107
[69]UK 2023 *Online Safety Act 2023* https://www.legislation.gov.uk/ukpga/2023/50/enacted, section 104
[70]Which appeared in the UK 2023 *Online Safety Act 2023* https://www.legislation.gov.uk/ukpga/2023/50/enacted, section 101, but requires amendments to the *Data Protection Act*. The process of implementing these amendments has paused during the UK election campaign, but both the Labour party and the Conservative party have made pledges to implement.

There are at least two key ways forward in the *Online Safety Act* with respect to more effective enforcement against overseas service providers. Both relate to the corporate structuring of foreign-owned digital platforms.

The first is amending the definitions of the Act so that overseas headquartered digital platforms are effectively 'onshored' and prevented from deflecting responsibility or liability to 're-seller' entities with no actual responsibility for or meaningful knowledge of the technical systems in use by the platform. This issue has been discussed at length in the *Andrew Forrest v Meta Platforms* litigation in the United States.[71] While not squarely in the context of the *Online Safety Act*, the litigation has revealed how vulnerable Australian enforcement actions are to sophisticated corporate structuring techniques.

The second point relates to insights gleaned from the recent case of *eSafety Commissioner v X Corp.*[72] The judgement offers a logic about online content that may challenge the future enforcement of the Act. His Honour accepted the Office of the eSafety Commissioner's decision that the content was class 1, meaning that it falls into the most severe category contemplated in the Act. Other content in this category includes material that shows or encourages child sexual abuse, for example.[73] There are clear and established policy reasons for why class 1 content should be removed from online services rather than merely geo-blocked, and the voluntary participation of other digital platforms to remove it either of their own accord or when asked informally by the Office of the eSafety Commissioner indicates the uncontroversial nature of class 1 material removal practices across the industry. His Honour held that it would be reasonable for X Corp to remove the content but unreasonable for the Office of the eSafety Commissioner to compel removal through section 109 of the Act.[74] Additionally, His Honour interpreted 'reasonable steps' in section 109 as demonstrably more modest than Office of the eSafety Commissioner's attempted action. The potential effect of this precedent is that any future attempts at enforcing removal notices under s 109 may be similarly rejected. To avoid future ambiguity and litigation risk, as well as improving safety standards, platforms that offer a service to Australian end-users must be required to implement 'reasonable steps' and similar requirements risk mitigation measures, as outlined in section 1.3 above.

Q20: 'Last resort' sanctions

Digital platforms do not have a strong track record of complying with Australian regulations where they lack enforcement powers. For example:

- Voluntary and co-regulatory requirements have been ignored. Such as, despite being a signatory to the voluntary *Australian Code on Disinformation and Misinformation* that places clear obligations on platforms to enable end-users to report misinformation,[75] X turned off the ability for Australians to report misinformation two weeks ahead of the Voice referendum.[76]

[71]United States District Court, California 2024 Andrew Forrest V. Meta Platforms Inc (2024) Case No. 22-cv-03699-PCP
[72]Federal Court of Australia, New South Wales Registry 2024 NSD474/2024 Esafety Commissioner V X Corp, FCA 499
[73]Office of the eSafety Commissioner 2021 *Online Content Scheme: Regulatory Guidance* https://www.esafety.gov.au/sites/default/files/2022-03/Online%20Content%20Scheme%20Regulatory%20Guidance.pdf
[74]Federal Court of Australia, New South Wales Registry 2024 NSD474/2024 Esafety Commissioner V X Corp, FCA 499 ([46])
[75]Digi 2022 *Australian Code on Disinformation and Misinformation* https://digi.org.au/wp-content/uploads/2022/12/Australian-Code-of-Practice-on-Disinformation-and-Misinformation-FINAL-_-December-22-2022.docx.pdf
[76]ABC News 2023 'Elon Musk's X reprimanded after disinformation safety feature scrapped' *ABC News* https://www.abc.net.au/news/2023-11-28/x-twitter-reprimanded-over-disinformation-safety-feature-removal/103158330

- Enforcement actions emerging from the *Online Safety Act*, such as requests for information from the Office of the eSafety Commissioner, have been ignored or inadequately responded to. To date, we believe three 'rounds' of non-periodic transparency reports have been requested by the Office of the eSafety Commissioner.[77] In response:

    - X did not comply with the non-periodic reporting notice regarding online hate by providing responses that were incorrect, significantly incomplete or irrelevant;[78]
    - Google failed to answer a number of questions in response to the notice regarding child sexual exploitation and abuse, and;[79]
    - X failed to provide any response to some questions, such as by leaving the boxes entirely blank, and in other instances, provided a response that was otherwise incomplete and/or inaccurate in response to the notice regarding child sexual exploitation and abuse.[80] The Office of the eSafety Commissioner issued X with an infringement notice for $610,500 in October 2023, which X challenged by seeking judicial review of the transparency notice, the service provider notification, and the infringement notice.[81]

- Regulations that do not explicitly designate platforms are 'ignored'. For example, Meta has announced an intention to walk away from deals negotiated within the framework of the *News Media Bargaining Code*.[82]

Changing this pattern of behaviour requires strong enforcement powers, including:

- Enhanced civil penalty powers, set to 10% of global annual turnover, and;
- Strong, 'last resort' measures. International best practice suggests that these last resort measures could include a range of:

    - Powers to temporarily restrict access to services (with judicial overview);[83]
    - Powers to limit monetisation within Australia (again with judicial overview),[84] and;
    - For breaches of transparency and information gathering requirements, criminal sanctions for identified senior managers.[85]

---

[77]Available at eSafety Commissioner 2024 *Responses to transparency notices* https://www.esafety.gov.au/industry/basic-online-safety-expectations/responses-to-transparency-notices
[78]eSafety Commissioner 2024 *Summary of response from X Corp. (Twitter) to eSafety's transparency notice on online hate* https://www.esafety.gov.au/sites/default/files/2024-01/Full-Report-Basic-Online-Safety-Expectations-Summary-of-response-from-X-CorpTwitter-to-eSafetys-transparency-notice-on-online%20hate.pdf, 19. Note "Subsequent information was provided by X Corp. after the Notice by the deadline that did seek to rectify earlier omissions of information provided. eSafety took this into consideration into account in deciding upon the appropriate enforcement action."
[79]eSafety Commissioner 2023 *Formal warning under section 58 of the Online Safety Act 2021 (Cth)* https://www.esafety.gov.au/sites/default/files/2023-10/Formal-Warning-to-Google-LLC.pdf
[80]eSafety Commissioner 2023 *Service provider notification in relation to contravention of section 57 of the Online Safety Act 2021 (Cth)* https://www.esafety.gov.au/sites/default/files/2023-10/Service-Provider-Notification-to-X-Corp.pdf eSafety Commissioner 2024 Responses to transparency notices https://www.esafety.gov.au/industry/basic-online-safety-expectations/responses-to-transparency-notices
[81]See *X Corp v eSafety Commissioner* (VID956/2023), status available at: https://www.comcourts.gov.au/file/Federal/P/VID956/2023/actions
[82]Georgia Roberts & Matthew Doran 2024 'Meta won't renew commercial deals with Australian news media' *ABC News* https://www.abc.net.au/news/2024-03-01/meta-won-t-renew-deal-with-australian-news-media/103533874
[83]Drawing from the EU experience, EU 2022 *Digital Services Act* https://eur-lex.europa.eu/eli/reg/2022/2065/oj, Article 51(3)
[84]Drawing from the UK experience, see UK Department for Science, Innovation & Technology 2024 *Online Safety Act: explainer* https://www.gov.uk/government/publications/online-safety-act-explainer/
[85]Drawing from the UK experience, see UK 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

Part 5: Questions 21-26

Q21: Incorporating international approaches

The EU's DSA provides a strong example of systemic, comprehensive and preventative legislation. The DSA sits alongside existing member state legislation around illegal content, and in a similar way in Australia could be implemented alongside the existing Complaints and content-based removal notices schemes to ensure complete coverage.

Likewise, the UK OSA provides a compelling framework for ensuring platforms' obligations towards end-users.

As we outline in section 1, we believe that lessons learned from these two models could be applied in Australia and see the BOSE replaced with a framework that:

1. Applies an overarching duty of care on platforms, from the UK OSA.

2. Mandates risk assessments based on the DSA that must consider at least the following risks:
   - The dissemination of illegal and harmful materials, as defined in the *Online Safety Act* Complaints and Contents-based Schemes;
   - The dissemination of online scams;
   - Negative effects on electoral processes and public security;
   - Negative effects to civil and political rights, such as political freedoms, and;
   - Negative effects on gender-based violence, children's best interests, public health and serious negative consequences to people's physical and mental wellbeing.

   Risk assessments must consider at least the following systems:
   - Recommender systems and other algorithms;
   - Content moderation systems;
   - Terms and conditions and their enforcement;
   - Systems for selection and presenting advertisements, and;
   - Data related practices of the provider where they create safety risks.

   Risk assessment templates could build on the ground breaking Safety by Design self-assessment tool developed by the Office of the eSafety Commissioner, but would need to be expanded and revised for this new purpose.

   Requirements for risk assessments could be 'tiered' and applied only to platforms that have significant numbers of Australian end-users.

3. Mandates for risk mitigation measures, learning from the UK OSA and DSA. Platforms must be required to implement adequate mitigation measures for each risk identified, that is commensurate to the scale and severity of the risk identified. Acceptable mitigation measures could include, but not be limited to
   - Changing design, features or functioning of their services, including their online interfaces;
   - Changing terms and conditions and their enforcement;
   - Changing content moderation processes;
   - Testing and changing algorithmic systems including recommender system;
   - Changing advertising systems, including the way ads are targeted at or presented to people;
   - Improving internal business processes to maximise safety;
   - Collaborating with other digital services;
   - Taking targeted measures to improve child safety, such as age assurance or parental control tools, and;
   - Ensuring evidence about potential illegal activities is stored and reported in helpful ways to law enforcement.

Requirements for risk mitigations could be 'tiered' and applied only to platforms that have to undertake risk assessments, i.e. have significant numbers of Australian end-users.

4. Enhanced measures for public transparency, building on the model in the DSA. Requirements for transparency could be 'tiered' and applied only to platforms that have significant numbers of Australian end-users.

5. Stronger enforcement regimes the ability to compel redress, so that regulators can ensure platforms change and improve safety standards. These could be backed by measures in the UK OSA and DSA:
    - Enhanced civil penalty powers, set to 10% of global annual turnover, and;
    - Strong, 'last resort' measures. International best practice suggests that these last resort measures could include a range of:
        - Powers to temporarily restrict access to services (with judicial overview);
        - Powers to limit monetisation within Australia (again with judicial overview), and;
        - For breaches of transparency and information gathering requirements, criminal sanctions for identified senior managers.

Q22: Statutory online duties

Incorporating a singular duty of care would help to 'flip the tables' and ensure accountability for user-safety rests with digital platforms rather than end-users. This approach has been used in the UK OSA, and to an extent, is also reflected in the Canadian Online Harms Bill. We note that the EU approach, of outlining a breadth of risks for which platforms have responsibilities may not translate adequately in the Australian context where we lack a broad 'Charter' to reference. Instead, a duty of care model has a long history in Commonwealth and Australian legislation, and offers significant opportunities to improve online safety.

As discussed in section 1, a duty of care is the first step in implementing a systemic approach to online safety regulation. A duty of care approach is systemic in that it moves the focus beyond the content layer to the underlying systems – the environment where content is created, shared and promoted. But as the UK Online Safety Network highlights, a 'systems focussed' approach extends beyond content by also covering content. A systems focus does not 'displace content rules. There are systems concerns here too. A service provider may have a policy prohibiting hate speech, but it might choose to run the platform in such a way that the policy is not enforced effectively: a weak system undermines the policy'.[86]

There is strong public support for these measures. Working with YouGov, in April 2024 we polled 1,514 people to gather their views on regulatory proposals.[87] Firstly, we asked about the proposals for regulation that focuses on content or systems, and found a strong preference for systemic regulation that works in conjunction with content focussed regulation (see Figure 6).

Laws that focus on risky content, so that risky content is taken down when it is found	*9%*
Laws that focus on systems, so that platforms are required to build in better and more effective ways to manage risky co	*20%*
Focus on both risky content and systems	*60%*
Neither	*3%*

[86]UK Online Safety Network 2024 *Submission to the Australian Department of Infrastructure, Transport, Regional Development, Communications and the Arts consultation regarding the Online Safety (Basic Online Safety Expectations) Determinations*, unpublished. Available on request
[87]Reset.Tech Australia 2024 *Digital Platform Regualtion*
https://au.reset.tech/news/green-paper-digital-platform-regulation/

Figure 6: Responses to the question 'There are a number of ways that laws can be made to try to improve online safety. Which of these would you prefer?' (n=1.514). 'Don't know' not included[88]

We also asked specifically about a duty of care, and found strong support for a duty of care, with 93% of people agreeing that social media companies should have a duty to take reasonable care of their users (see Figure 7).

Agree	*93%*
Disagree	*5%*

Figure 7: Responses to the question 'Social media companies should have a duty to take reasonable care of their user' (n=1,514). 'Don't know' not included[89]

Implementing a duty of care in Australia creates obligations to address risks that the systems and processes of digital platforms create. A duty of care model includes four aspects:

- 'The overarching obligation to exercise care in relation to user harm;
- Risk assessment process;
- Establishment of mitigating measures, and;
- Ongoing assessment of the effectiveness of the measures.'[90]

Introducing an overarching duty of care model in the *Online Safety A*ct would comfortably sit alongside the Complaints and content-based removal notices schemes, but could transform safety expectations into a systemic, comprehensive, preventive approach.

We note that while industry may have a preference for implementing specific duties of care regarding specific types of content, as noted in section 1.1 this:

- Creates "gaps" in protections for end-users. Duties tied to particular sort of content cannot address harms arising from overarching abusive designs that do not fall into a, such as dark patterns that deceive users or abusive design techniques deployed at children;

- Hampers the systemic and preventative approach. A singular duty of care approach acknowledges that, for digital platforms, systems are developed and business decisions are made *before* such platforms are actually populated with content. Platforms decide how their content recommender systems will work, or how their moderation teams will be staffed etc., without knowing what content they will recommend or moderate each day. A singular duty of care approach encourages platforms to safeguard these systems before any harm has happened and before any designated content has been posted. However, implementing duties of care tied to particular sorts of content requires platforms to risk assess their systems after they are 'populated' with designated content. This seems at odds with the sort of "upstream" and preventative approach that a duty of care seeks to enable, and;

- Moves the regulation away from a focus on the systems and back into specifying particular types of content. This skews the focus of compliance towards a content-first rather than a systems-first approach. This was present in much of the Parliamentary debate in the UK,

[88]Reset.Tech Australia 2024 *Digital Platform Regualtion* https://au.reset.tech/news/green-paper-digital-platform-regulation/
[89]Reset.Tech Australia 2024 *Digital Platform Regualtion* https://au.reset.tech/news/green-paper-digital-platform-regulation/
[90]UK Online Safety Network 2024 *Submission to the Australian Department of Infrastructure, Transport, Regional Development, Communications and the Arts consultation regarding the Online Safety (Basic Online Safety Expectations) Determinations*, unpublished. Available on request

which became very focused on what content would be removed and what would not as duties around content type emerged.[91]

Q23: Current levels of transparency

Improving public transparency enhances safety and trust. Building on international example, Australia could adopt a model of transparency that includes requirements for:

- Summaries of risk assessment being published after a reasonable period of time;
- Annual transparency reports, which are detailed and prescriptive (see figure 3 for examples of potential prescriptions);
- Annual independent audits. An independent expert or 'skilled person' should be required to review both the platforms' risk assessment and transparency notice;
- Data portals. This would include searchable ad repositories, and data about content moderation decisions, and;
- Researcher access requirements. Vetted researchers in Australia should be able to request data (see section 1.4 for more details).

There is strong public support for these measures. Working with YouGov, in April 2024 we polled 1,514 people to gather their views on regulatory proposals.[92] We asked about transparency and found strong public support for a battery of transparency measures (see Figure 8).

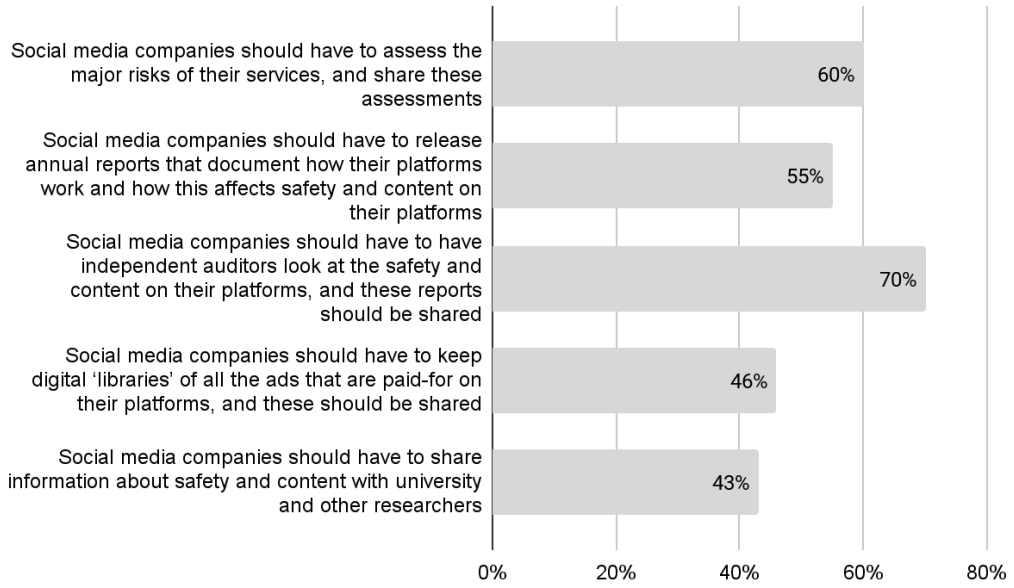

Figure 8: Responses to the question: 'It's not always clear how social media companies build their systems and algorithms. There are some discussions that laws could be passed that make social media companies be more transparent about how platforms work and the consequences of this. Which, if any, of these transparency measures would you support in law?' (select all you support)' (n=1,514)[93]

[91]See Lorna Woods & Rys Farthing 2024 'The dangers of pluralisation: A singular duty of care in the Online Safety Act' *The Policy Maker* https://thepolicymaker.jmi.org.au/the-dangers-of-pluralisation-a-singular-duty-of-care-in-the-online-safety-act/

[92]Reset.Tech Australia 2024 *Digital Platform Regualtion* https://au.reset.tech/news/green-paper-digital-platform-regulation/

[93]Reset.Tech Australia 2024 *Digital Platform Regualtion* https://au.reset.tech/news/green-paper-digital-platform-regulation/

Public transparency requirements could be 'tiered' and applied only to platforms that have significant numbers of Australian end-users.

Q24: Researcher and regulator access to data

Vetted researchers in Australia should be able to request data. Like in the EU, requirements for an Australian vetted researcher could include:

- Affiliation to a research organisation, including academic and third sector research organisations;
- Researchers or at least the lead researcher should be an Australian resident, and;
- Non-commercial purpose limitations;

Suitable research projects are provided with data. A suitable proposal would include information demonstrating that:

- The research fits the ambitions of the *Online Safety Act* and how it is broadly of public benefit. This does not include data about trade secrets;
- Funding for the research is fully disclosed;
- Access to the specific data requested, and the indicated timeline indicated, is necessary and proportionate to the purposes of the research;
- Data security and confidentiality requirements, as well as personal data safety requirements, will be fulfilled, and;
- The research results will be made publicly available free of charge within a reasonable period after the completion of the research.

The process for researchers requesting data could be managed by the ACMA, the Office of the eSafety Commissioner or other appointed independent organisation. In addition, existing data and data tools like APIs should be made available to Australian researchers for free. Currently for example, the TikTok researcher API is only available to European and American researchers.

Australian regulators should also have access to this data.

Q25: Ombuds scheme

There are limited effective dispute resolution systems for 'systemic' risks, and also for end-users whose harm falls outside the scope of the existing Complaints and content-based removal notices schemes. An expanded complaints scheme, or an ombuds scheme could be welcome to address these gaps. However, the difference between and interactions between an Ombuds scheme and the existing, excellent public facing complaints mechanism run by the Office of the eSafety Commissioner would need to be clarified.

As we have mentioned above, the UK OSA creates a scheme for super complaints to be made to the regulator that might be useful for consideration in this regard.[94]

Q26: Additional safeguards to protect human rights

The review of the *Online Safety Act* presents an opportunity to advance children's rights within the digital world. As the submission from the Australian Child Rights Taskforce notes, children's rights in relation to the digital environment are comprehensive. There is an emerging

[94]UK Department for Science and Innovation 2023 *Children's charities and free speech groups could be allowed to submit super-complaints to Ofcom to keep internet safe* https://www.gov.uk/government/news/childrens-charities-and-free-speech-groups-could-be-allowed-to-submit-super-complaints-to-ofcom-to-keep-internet-safe and UK Department for Science and Innovation 2023 *Super-complaints: eligible entity criteria and procedural requirements* https://www.gov.uk/government/consultations/super-complaints-eligible-entity-criteria-and-procedural-requirements

regulatory trend towards advancing children's rights in digital regulation by introducing the 'children's best interests principle' into regulation that affects the digital world. This includes international regulations, such as the UK's *Age Appropriate Design Code* which takes a rights-based approach to data protection for children,[95] and also in Australia. For example, proposals for reform to the *Privacy Act*[96] include options such as:

- Requirements to consider children's best interests in deciding if data processings is 'fair and reasonable';
- The introduction of a Children's Privacy Code, which would embeds the best interest principle, and;
- Requirements prohibiting direct marketing to children under 18 years and prohibiting targeting children under 18 years except where it is in their best interests.

The introduction of requirements to ensure industry acts in the best interests of the child in the *Online Safety Act* presents a significant opportunity to advance children's rights in their own right, but also may help to 'join up' privacy and online safety protections for children in particular creating comprehensive, rights focussed protections.

We note that determining children's best interests is not always straightforward, and clear guidance around this could be helpful.[97]

Introducing the best interests principle into the *Online Safety Act* enjoys broad public support. In April 2024 we commissioned YouGov to poll 1,515 adults about proposals to introduce the children's best interests principle into privacy and safety laws. Fifteen percent of respondents thought the children's best interest principle should be in place to protect the use of children's data alone (privacy), 12% thought it should be in place when it came to online safety rules alone and 67% thought it should be in place for both. In total, 79% of people thought the best interest principle should be included in online safety frameworks (see Figure 9).

[95]UK ICO 2020 *Age Appropriate Design Code*
https://ico.org.uk/for-organisations/uk-gdpr-guidance-and-resources/childrens-information/childrens-code-guidance-and-resources/age-appropriate-design-a-code-of-practice-for-online-services/
[96]Attorney General's Department 2023 *Privacy Act Review Report,*
https://www.ag.gov.au/rights-and-protections/publications/privacy-act-review-report
[97]See for example, a first attempt at what this might look like in the privacy domain at Reset.Tech Australia 2024 *Best Interets and Targeting: Implementing the Privacy Act Review to advance children's rights*
https://au.reset.tech/news/best-interests-and-targeting-implementing-the-privacy-act-review-to-advance-children-s-rights/

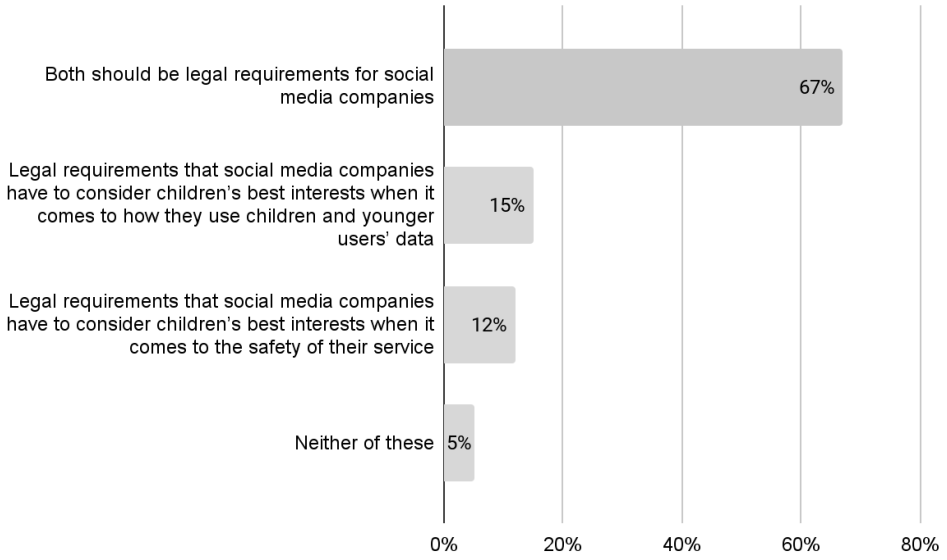

Figure 9: Responses to the question 'It's not always clear if social media companies make their products in ways that are best for children and younger users under 18. There are some discussions that laws could be passed that make social media companies think about children's best interests in the way they work. Which, if any, of these measures would you support in law?' (n=1,514). 'Don't know' not included[98]

Part 6: Questions 27-33

The role of platform's systems and processes in both generating and amplifying all of the risks noted in section 6 is significant. A systemically focussed bill, which implements an overarching duty of care on platforms which is realised through risk assessments and risk, should mitigate against a full breadth of harms in an 'upstream' way. We reiterate our concerns that in adding long lists of harmful content-types or conduct-types runs the risk of the *Online Safety Act* becoming a 'Christmas tree bill'. Overly embellished legislation is not future proofed aslists of harms become out-of-date almost immediately. Further, this creates the space for contested definitional issues as new harms emerge.

Q27: Group or individual harms

A focus on individual harms is narrow and leaves Australians vulnerable to collective risks. Collective risks come in two interconnected forms:

1. Group or community risks, such as indigenous communities, CALD communities, women and LGBTIQ+ people. These communities often suffer unique and disproportionate harms in the digital world. While some of the risks they face may be addressed by regulation around individual harms, an 'offensive-piece-of-content' by 'offensive-piece-of-content' approach can miss the collective nature of the problem.

2. Societal risks. The scale and reach of social media platforms has the capacity to influence and affect Australian institutions, such as Parliament, the Press and healthcare systems, often with destabilising effects. For example, we have seen how digital platforms have

[98]Reset.Tech Australia 2024 *Digital Platform Regualtion*
https://au.reset.tech/news/green-paper-digital-platform-regulation/

been used to undermine public health messaging around vaccine roll out (often in ways with particular consequences for marginalised communities), and foreign bots engaged in Australian electoral discussions.[99]

A systemic focus, made comprehensive through specific risk assessment requirements as outlined in section 1.2, would support minimising risks that affect groups, as well as societal level risks.

Q31: What features of the Act are working well

The public facing complaints mechanism in the Complaints and content-based removal notices schemes are world leading, and for those who have been harmed in specific ways as covered by the Act, it can be life changing.

As we have noted, access to this scheme is limited to those affected by specific types of content and precludes complaints based on systemic issues. We believe that within a comprehensive framework, this system could be expanded to protect individuals affected by a border range of risks and to allow super-complaints to be made around systemic risks. The latter might form part of an ombuds scheme, but it is not clear to us how these would differ and interact at this stage.

Q33: Cost recovery mechanisms

The role of the Office of the eSafety Commissioner is absolutely central to reducing the online risks faced by Australians. However, comparatively, this office is under-resourced. Ofcom anticipates recruiting around 300 full time staff to implement the UK OSA (in addition to their existing staff of 1,000).[100] At the European Commission, we know at least 40 jobs in enforcement will be created (at DG Connect)[101] as well as 30 at the Centre for Algorithmic Transparency,[102] not including the engagement team or country level Digital Service Coordinators. Australia is a much smaller market and staffing needs will reflect this, but enhancing resources at the Office of eSafety Commissioner would help. Cost recovery mechanisms make sense as a way to ensure that the taxpayer is not meeting this burden.

[99]Felicity Caldwell 2019 'Bots stormed Twitter in theirthousands during the federal election' *Sydney Morning Herald*
https://www.smh.com.au/politics/federal/bots-stormed-twitter-in-their-thousands-during-the-federal-election-20190719-p528s0.html
[100]Ofcom 2021 *Ofcom's perspective on draft online safety legislation: Letter to Julian Knight & Damian Collins Digital, Culture, Media and Sport Committee*
https://www.ofcom.org.uk/siteassets/resources/documents/about-ofcom/public-correspondence/2021/letter-melanie-dawes-draft-online-safety-legislation.pdf
[101]European Commission 2024 *Job opportunities within the Digital Services Act Enforcement Team*
https://eu-careers.europa.eu/sites/default/files/eu_vacancies/2024-01/2024_2nd%20call_Job_opportunities%20within%20the%20DSA%20Enforcement%20Team%20final%20with%20privacy%20statement%20Final.pdf
[102]Théophane Hartmann 2024 'Challenges mount for European Commission's new DSA enforcement team' *Euroactive*
https://www.euractiv.com/section/law-enforcement/news/challenges-mount-for-european-commissions-new-dsa-enforcement-team/