Reset.
AUSTRALIA

The future of digital regulation in Australia:

Five policy principles for a safer digital world

Reset Australia
May 2022

Executive Summary

As Australia implements much needed digital regulation on a number of issues, there is a need to ensure policy initiatives are joined up and coherent. To avoid creating inconsistencies and gaps, and to ensure an effective approach to reshaping our digital regulatory landscape, this document presents five overarching policy principles that could function as a 'North Star' for policy makers.

These principles include:

1. **Systems and processes:** Focus regulation on eliminating risks from systems and processes, expanding on our current focus on content moderation
2. **Community and societal risk:** Expand regulations to addresses community & societal risks, building on our comprehensive approach to Individual risks
3. **Platform accountability and transparency:** Ensure regulation creates accountability & transparency, rather than placing burden on individuals
4. **Comprehensive regulation:** Ensure the regulatory framework is comprehensive, by improving our current regulatory gaps and disjunctures
5. **Strong regulators and enforced regulation:** Ensure regulation is strong and enforced, by moving away from self- and co-regulation and resourcing and joining up regulators

This would create a more streamlined approach to regulation, replacing multiple disjointed obligations with more aligned upstream duties, and reducing the regulatory burden on Australia's successful tech industry.

This approach would also be interoperable with emerging international requirements, ensuring Australian industry could expand into international markets with minimal regulatory friction.

Contents

Introduction

Australia is moving swiftly to implement critical significant regulatory reforms in the digital world, exploring responses to online safety, defamation, monopoly power and privacy. It's easy to get disoriented in this whirlwind of policy proliferation, which may lead to disjointed or inconsistent regulations being implement. This document presents Reset's vision of a 'North Star', in the hope that it is helpful for policy makers and civil society as Australia develops and refreshes its digital regulatory landscape.

An updated, stronger regulatory landscape is needed now more than ever. When powerful forces 'move fast and break things', they can leave a wake of destruction in their path. Social media platforms and other Big Tech companies have often done exactly that in the past. Australian institutions, including Parliament and the press, have been exposed to harmful mis/disinformation and arbitrary shut downs. Individual Australians have had their data illegally harvested, and their privacy invaded. The porousness of the regulatory landscape, combined with the lack of precaution demonstrated by Big Tech, has too often destabilised Australian communities, with serious online and offline consequences.

Yet, this rapid change has also created a digital world ripe with opportunities, generating innovations that strengthen the economy and improve our lives. In 2021, the Australian tech sector contributed $167bn to the economy[1], and during the pandemic kept many families and children connected, working and learning. The digital world can be a force for good, and the impact of the technology sectors, if directed and regulated, can be transformative.

What Australia needs is a citizen-focussed, strong, coherent regulatory landscape that harnesses the potentials of the digital world for good. We believe that five overarching policy principles are needed to reshape our regulatory framework, and to ultimately protect all Australians whilst continuing the growth of a vibrant Australian tech sector.

Reset Australia is an independent, non-partisan policy think tank 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] Tech Council & Accenture 2020 *The Economic Contributions of Australia's Tech Sector* https://techcouncil.com.au/wp-content/uploads/2021/08/TCA-Tech-sectors-economic-contribution-full-res.pdf

1. Eliminating risks from systems and processes

Australia regulation should pivot towards targeting risks created across the systems and processes developed by digital services. The aspects of systems and processes, and related risks, that regulation could address includes:

- **Algorithms.** These drive much of the content delivery in social media platforms, both in terms of content and advertising. For example, YouTube estimates that 70% of content viewed on their platform is as a result of their recommender algorithm and autoplay. These systems, designed by platforms, often using machine learning or other AI technologies, often promote risky or harmful content. Yet algorithms are not trained in ways that consider risks.

- **Platform design.** The user interface and user experiences of social media platforms are highly curated and engineered: each design element reflects a decision point made by a company. Platforms can be designed in ways that create risks. For example, many platforms design their user journey in ways that maximise data extraction from your device, social media apps and other internet based activity. For example, apps that nudge you to, or automatically connect with, your address book or track your GPS location. This data is used to preferences and interests, and personalise your ad experience (termed 'surveillance advertising'). This is a 'dark pattern' that maximises profits but does not consider the data risks it creates, in subtle and persuasive ways.

- **Specific features.** Specific features can also create risks. For example, features that enable the live broadcasting of locations, or photo filters that make people appear thinner. These features can combine in ways that amplify or create new risks. For example, video live streaming and the ability to receive messages from stranger's accounts creates unique risks for young users[2]. Features are developed and refined by platforms to meet identified priorities, such as maximising engagement, growing reach or extending the amount of time users stay on a platform. These priorities often do not consider risk; if 'minimising risk' was a systemic design aim many features would operate differently or be abandoned.

- **Age assurance processes.** Platforms can use a range of systems and techniques to estimate or validate the age of their users. Knowing the age of users can be useful for a range of purposes, from preventing access to age restricted content to providing a more safe and secure user experience for children. Clarity around the *intent* of the assurance is needed, to ensure that it can be mapped to the most appropriate *method*. A blunt approach to implementing age assurance can create unnecessary privacy risks and cause unintended harm, including to young users.

[2]See for example, The Times' investigation in grooming via YouTube livestreams. Harry Shukman 2018 'Predators coax children into exposing themselves' *The Times* www.thetimes.co.uk/article/predators-coax-children-into-exposing-themselves-lfws0fjdp

These sorts of systems and processes manufacture and amplify risks but none of them are inevitable. Social media platforms can change and improve their systems, and regulation can encourage them to do so.

Regulatory approaches that take a more narrow focus on content moderation (focusing on takedown/deletion of harmful or illegal content for example) are not systemic enough, nor are they commensurate with the scale of the problem at hand. They doom regulators to a perpetual game of content 'whack-a-mole' on an impossible scale.

Australia's existing *Online Safety Act* focuses largely on content, but through the Basic Online Safety Expectations and the industry codes developed as part of this may address some systemic risks. However, co-regulatory codes and guidance from regulators will not be adequate to create the scale of change needed to ensure safety. These risks are simply too important to leave up to industry to address — whose business models incentivise and reward risky systems. Nor will the proposed Codes cover all of the systems and processes that need to be addressed. A more comprehensive approach is needed to ensure the regulatory framework is fit for newly emerged and emerging technologies.

CASE STUDY: HOW A RISK FOCUSED, SYSTEMIC APPROACH WORKED TO PROTECT CHILDREN

Regulations that remove risks from systems have already reduced risks for children and young people. Without regulating content, the UK's *Age Appropriate Design code* led to 'upstream' risk reductions such as:

- Defaulting children's accounts to private. In the 8 months leading up to the enforcement of the UK's code, TikTok announced that it was defaulting all users aged 13-15 to private accounts[3], Facebook announced that 'everyone who is under 16 years old (or under 18 in certain countries) will be defaulted into a private account when they join Instagram[4]' and Google announced that it would 'gradually start adjusting the default upload setting to the most private option available for users ages 13-17 on YouTube[5]'
- Reducing the ways advertisers themselves could micro-target commercial advertising at children. Google announced it was blocking microtargeting based on age, gender or interests of people under 18[6], and Meta limit the ability of advertisers to select children to

[3] Eric Han 2021 'Strengthening privacy and safety for youth on TikTok'
newsroom.tiktok.com/en-us/strengthening-privacy-and-safety-for-youth
[4] Instagram 2021 'Giving young people a safer, more private experience'
about.instagram.com/blog/announcements/giving-young-people-a-safer-more-private-experience
[5] James Beser 2021 'New safety and digital wellbeing options for younger people on YouTube''
blog.youtube/news-and-events/new-safety-and-digital-wellbeing-options-younger-people-youtube-and-youtube-kid
[6] James Beser 2021 'New safety and digital wellbeing options for younger people on YouTube'
blog.youtube/news-and-events/new-safety-and-digital-wellbeing-options-younger-people-youtube-and-youtube-kid

target, allowing selected targeting based on age, gender and geography[7], and in-platform tracking to personalise ads[8]

- Turning off 'Autoplay' by default features which can see children 'nudged' into watching more content than they intended. For example, Google and subsidiary YouTube announced they would turn off Autoplay for those under 18[9]

Appendix 1 outlines what this approach may look like for children more holistically.

ELIMINATING RISKS FROM SYSTEMS & PROCESSES: EXAMPLES FROM OTHER JURISDICTIONS

The UK government described their draft Online Safety Act as 'a "systems and processes" bill — aimed at addressing systemic issues with online platforms rather than seeking to regulate individual content'. The final act is expected to focus on the 'content and activity' of platforms[10]. The bill achieves this by creating multiple and overlapping 'duties of care' for service providers, including[11]:

- Duties to reduce illegal content risks, such as:
  - Undertake an illegal content risk assessment
  - Taking proportionate steps to mitigate and effectively manage risks identified in illegal content risk assessments
  - A duty to operate using proportionate systems and processes designed to minimise the presence, duration of presence and dissemination of illegal content
- Duties to regard freedom of expression and privacy set, such as:
  - Impact assessment about free expression & privacy, when deciding safety policies & procedures
  - Likewise, impact assess existing policies & procedures
  - Duties to act on risks identified in these assessments
- Duties about reporting and redress, with similar obligations
- Record-keeping and review duties, with similar obligations

---

[7] Instagram 2021 'Giving young people a safer, more private experience' about.instagram.com/blog/announcements/giving-young-people-a-safer-more-private-experience

[8] Elena Yi-Ching Ho & Rys Farthing 2021 *How Facebook are still targeting teens with advertising* Reset Australia / Fairplay https://fairplayforkids.org/wp-content/uploads/2021/11/fbsurveillancereport.pdf

[9] James Beser 2021 'New safety and digital wellbeing options for younger people on YouTube' blog.youtube/news-and-events/new-safety-and-digital-wellbeing-options-younger-people-youtube-and-youtube-kid

[10] Joint Committee on the Draft Online Safety Bill 2021 *Draft Online Safety Bill: Report of Sessions 2021-22* https://committees.parliament.uk/publications/8206/documents/84092/default/

[11] *Draft Online Safety Bill* 2021, UK https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/

In a similar vein, the EU's proposed Digital Services Act (DSA) will oblige platforms to[12]:

> *'assess the systemic risks stemming from the functioning and use of their service, as well as by potential misuses by the recipients of the service, and take appropriate mitigating measures'*

Article 58 of the DSA goes on place requirements on platforms to diligently mitigate risks identified in the risk assessment, by for example:

> *'enhancing or otherwise adapting the design and functioning of their content moderation, algorithmic recommender systems and online interfaces, so that they discourage and limit the dissemination of illegal content, adapting their decision-making processes, or adapting their terms and conditions. They may also include corrective measures, such as discontinuing advertising revenue for specific content, or other actions, such as improving the visibility of authoritative information sources. … They may also initiate or increase cooperation with trusted flaggers….'*

2. Expand regulations to address community & societal risks

The risks addressed by existing legislation are too narrow, and this leaves Australians vulnerable to collective risks. Collective risks come in two interconnected forms.

Firstly, there are risks posed to specific communities, such as indigenous communities, migrant communities, people of colour, 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. Disinformation and hate speech can affect particular communities in ways that differ from individual harm.

Secondly, platforms create 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 social media platforms have 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. This is not the stuff of 'conspiracy theories'; a 2021 Senate hearing revealed that Australia has been the target of a number of sophisticated foreign disinformation campaigns, including a network

[12] Recital 56 European Commission 2020 *Proposal For A Regulation of the European Parliament & of the Council on a Single Market For Digital Services (Digital Services Act) & Amending Directive* 2000/31/ https://oeil.secure.europarl.europa.eu/oeil/popups/ficheprocedure.do?lang=en&reference=2020/0361

linked to marketing firms based in the UAE, Nigeria and Egypt, all enabled by platforms[13]. A further list of examples of electoral and other disinformation in Australia is provided in appendix 2.

Expanding the definitions of harms (and risks) addressed in Australia's regulatory framework would better protect Australian communities and society at large. This means tackling mis and disinformation, and explicitly addressing hate speech. Currently mis/disinformation is covered by a co-regulatory Code that has been widely criticised as 'not meeting expectations' including by regulators[14].

The focus on reducing community and societal harms must also take a systems and processes focus. For example:

- Addressing the potential societal harms caused by mis and dis information requires a critical focus on advertising recommender systems. Advertising algorithms that focus entirely on maximising engagement provide equal access to financial incentives to both good actors to post great content, and to bad actors to post harmful content. Any content that gets traffic is monetizable through undiscerning advertising algorithms. This business model has been shown to contribute substantially to the funding model of disinformation[15] and hate speech[16].

- Addressing the societal harms of divisive content requires a focus on content recommender systems. Facebook's use of 'engagement based ranking' to prioritise content in its algorithm, for example, amplifies harm. Engagement (be it a click through, a comment or a reshare) drives profits for platforms. However content which elicits an extreme reaction, be it inflammatory divisive[17] or misinformation[18], is more likely to encourage engagement. This means divisive content and disinformation are systematically over-promoted in people's feeds. This can have consequences for political discourse, fragmenting and polarising society, and hindering our capacity for genuine political conversations. This threat is particularly salient during elections when manipulation of the information ecosystem can sway votes.

[13] Select Committee on Foreign Interference Through Social Media, Senate, 30 July 2021
[14] Zoe Samios & Lisa Visentin 2020 'ACMA: Tech giants' code to handle fake news fails to meet expectations' *SMH* www.smh.com.au/politics/federal/acma-tech-giants-code-to-handle-fake-news-fails-to-meet-expectations-20201026-p568oq.html
[15] Global Disinformation Index 2020 *Ad-funded Covid 19 disinformation* https://disinformationindex.org/wp-content/uploads/2020/07/GDI_Ad-funded-COVID-19-Disinformation-1.pdf
[16] Karen Hao 2021 'How Facebook and Google Fund Global Misinformation' *MIT Technology Review* www.technologyreview.com/2021/11/20/1039076/facebook-google-disinformation-clickbait/
[17] Luke Munn 2020 'Angry by design: toxic communication and technical architectures' *Humanities and Social Sciences Communications* doi.org/10.1057/s41599-020-00550-7
[18] Peter Dizikes 2018 'On Twitter, false news travels faster than true stories' *MIT News* https://news.mit.edu/2018/study-twitter-false-news-travels-faster-true-stories-0308

- Content recommender systems have been shown to consistently suffer from issues around racism[19] and sexism[20]. Likewise, search engines have been shown to amplify race and sex descrimnination[21]. A recent experiment in Australia found that it took TikTok's recommender algorithm only 7 hour and 42 minutes to 'learn' that an account was interested in content that promoted harmful gender stereotypes and began to recommend this content at such a frequency that it would take only 5-6 days of regular use before their social media feed was completely filled with this content[22].

CASE STUDY: SOCIETAL RISKS IN ELECTION PROCESSES

A QUT study which examined around 54,000 accounts during and after the 2019 Australian Federal Election (looking at over 1 million tweets) revealed that 13% of accounts were 'very likely' to be bots, with the majority originating from New York[23]. This is estimated to be more than double the rate of bot accounts in the US presidential election.

These can have big impacts: research into the US election by ANU indicated that the average bot was 2.5 times more influential than the average human, measured by success at attracting exposure via retweets[24].

CASE STUDY: SOCIETAL AND COMMUNITY RISKS THROUGH MIS/DISINFORMATION

Chinese Australians have faced misinformation in the past, often in what appear to be coordinated disinformation campaigns[25]. Social media platforms, such as WeChat, Weibo and Douyin have been found to serve targeted misinformation to Chinese language speakers in Australia. In 2019,

---

[19] Derek O'Callaghan, Derek Greene, Maura Conway, Joe Carthy, Pádraig Cunningham 2014 'Down the (White) Rabbit Hole: The Extreme Right and Online Recommender Systems' *Social Science Computer Review* doi.org/10.1177/0894439314555329

[20] Masoud Mansoury, Himan Abdollahpouri, Jessie Smith *et al* 2020 'Investigating Potential Factors Associated with Gender Discrimination in Collaborative Recommender Systems' *Cornell University Computer Science* *arXiv:2002.07786*

[21] Safia Noble 2018 *Algorithms of Oppression* NYU Press

[22] Dylan Williams, Alex McIntosh & Rys Farthing 2021 *Surveilling young people online* Reset Australia au.reset.tech/uploads/resettechaustralia_policymemo_tiktok_final_online.pdf

[23] Felicity Caldwell 2019 *'Bots stormed Twitter in their thousands during the federal election'* SMH www.smh.com.au/politics/federal/bots-stormed-twitter-in-their-thousands-during-the-federal-election-20190719-p528s0.html

[24] Sherryn Groch 2018 '*Twitter bots more influential than people in US election: research'* SMH www.smh.com.au/national/twitter-bots-more-influential-than-people-in-us-election-research-20180913

[25] Lawson 2020 'WeChat the channel for China disinformation campaigns' *Canberra Times* www.canberratimes.com

WeChat in particular was a site of much political campaigning in Mandarin which included mis & disinformation[26].

ADDRESSING COMMUNITY & SOCIETAL RISKS: EXAMPLES FROM OTHER JURISDICTIONS

The EU and Germany are moving towards frameworks that address community and societal risks, as well as individual risks. Figure one documents these.

Recital 57 of the Digital Services Act explicitly describes the three types of systemic risks that platforms must assess and mitigate, which include community and societal risks[27]:

1. Risks associated with the misuse of their service through the dissemination of illegal content, such as CSAM, hate speech, and the conduct of illegal activities. This includes risks created where content may be amplified by platforms to an especially vast audience. These are largely, but not exclusively, individual risks
2. Risks that affect people's rights. This includes the design of the algorithmic systems and the misuse of their service through the submission of abusive notices or other methods for silencing speech or hampering competition. These would include community risks as we have described them
3. The use of a platform to share disinformation that has a foreseeable impact on health, civic discourse, electoral processes, public security and children's safety. This includes mitigating against fake accounts and bots. These would include societal risks as we have described them

[26] Lawson 2020 'WeChat the channel for China disinformation campaigns' *Canberra Times* www.canberratimes.com

[27] Recital 57, European Commission 2020 *Proposal For A Regulation of the European Parliament & of the Council on a Single Market For Digital Services (Digital Services Act) & Amending Directive* 2000/31/ https://oeil.secure.europarl.europa.eu/oeil/popups/ficheprocedure.do?lang=en&reference=2020/0361(COD)

	EU	CANADA	GERMANY	UK	IRELAND	AUSTRALIA
Key legislation addressing harms	Digital Services Act (in draft)	Online safety proposals (currently being redrafted)	NetzDG, and others (in force)	Online Safety Bill (in draft)	Online safety & Media Regulation (in draft)	Online Safety Act (in force)
Definition of Harm, Individual, Community or Societal	No set definition, the focus is on harms that violate rights. This will include societal harms, and community harm through hate speech	Individual (aligned to existing definitions of hate speech) Societal (damage to societal cohesion, vulnerable groups)	Based on existing criminal law. This includes Individual and some community harms through hate speech	Individual (Content having an adverse physical or psychological response on adults of children)	Individual (Illegal content, individually intimidating or threatening content, eating disorder, self harm & suicide content)	Individual (content that is "offensive" to adults or children, content that is refused classification etc)
Systems Vs Takedown	Systems + Takedown	Takedown	Takedown	Systems + Takedown	Systems + Takedown	Takedown (+ potentially some systems through co-regulatory Codes)
Content In Scope	Illegal + indirectly, legal Disinfo included indirectly Hate speech indirectly included	Illegal Disinfo out of scope Hate speech in scope	Illegal Disinfo out of scope Hate speech in scope	Illegal + legal List of harms to be added later but unclear whether disinfo & hate speech is in scope (could be in scope where content is harmful to adults)	Illegal + legal Disinfo out of scope Individual hate speech content could be in scope, where it intimidates, threatens, humiliates or persecutes	Illegal + legal Disinfo out of scope Individual hate speech content could be in scope, where it causes offence to an individual or would be considered menacing, harassing or offensive
Services In Scope	Intermediary services e.g. ISPs and online platforms Private messaging out of scope	Social media Private messaging out of scope	Social media	Services which host or facilitate UGC, apart from news media outlets. Private messaging in scope.	Broad range of platforms and services inc press publications which enable UGC Private messaging in for criminal content	Social media services, Relevant electronic service and ISPs (Tight definition of "social media")
Powers Of Regulator	Fines	Information gathering powers	Fines	Fines	Fines	Fines

	Information gathering powers Algorithmic audit mandatory	Inspection powers No algorithmic audit		Information gathering powers Language seems to allow algorithmic inspection	Information gathering powers. No algorithmic audit	Offers public facing complaint mechanisms, Investigation, Audit (not algorithmic)
Independence Of Regulator	Independent as well as EC oversight of large platforms	Independent Creates Digital Safety Commissioner and Digital Recourse Council of Canada,	Independent	Independent however OSB keeps provisions for political agenda setting	Independent Creates Online Safety Commissioners	Independent
Transparency	Six monthly transparency reports (publicly published) Data access for pre-vetted researchers	Transparency reporting inc data on takedown volumes and processes.		Annual transparency reports No data sharing provisions	Periodic transparency reporting	Transparency reporting

Figure one: Comparative approaches to types addressing harms through regulation

3. Ensure regulation creates accountability & transparency

There are multiple ways governments can regulate the digital world, but the most effective policies require accountability and transparency from tech platforms themselves. Regulations that identify the core risks as stemming from platforms themselves — and squarely place the burden of responsibility on digital services — should be prioritised.

Regulation can place duties on users in multiple ways, but these are often inappropriate or ineffective:

- Solutions that position individual users (especially children and parents) as key actors in the frontline of improving safety are often inappropriate and will fail to protect all Australians. The scale of the risks created by platforms exceed the capability of individuals to effectively manage in isolation, especially for children. The ability to 'change settings', 'effectively report content' or 'turn on safe search' will not be enough. User's informed choice around settings and options is necessary, but it is not sufficient to ensure safety, particularly for those lacking the capabilities or support to do so

- Solutions that pass responsibility on to users (as parents or consumers) to read 'the fine print' or consent to a risky system misrepresents the power asymmetry between users and tech companies. The nature of the global digital architecture, and its utility in everyday life, means that withdrawing consent is not a viable option for most Australians. For example, 75% of the world's most popular million websites have google analytics and trackers built into them[28]. A 'buyer beware' approach will fail where users have no viable alternatives

- Solutions that position individual users (be they 'trolls' or influencers) as the key actors responsible for harm undersells the role of platforms in creating the risky digital environments that enable and encourage toxic actors. Platforms manufacture and amplify harmful content; they hand trolls and other bad actors the tools they need to cause harm and provide incentives, including funding[29], to encourage their ongoing poor behaviour

Accountability means that platforms themselves should have responsibilities to mitigate risks, and should be held to account where they fail and harm occurs. This upstream focus in keeping with existing norms around effective ways to reduce industrial hazards. The hierarchy of hazard controls — a globally used framework — outlines that the most effect interventions emerge from eliminating hazards[30] (see

[28] Steven Englehardt & Arvind Narayanan 2016 'Online Tracking: A 1-million-site Measurement & Analysis' *CCS '16: Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security* doi.org/10.1145/2976749.2978313

[29] Karen Hao 2021 'How Facebook and Google Fund Global Misinformation' *MIT Technology Review* www.technologyreview.com/2021/11/20/1039076/facebook-google-disinformation-clickbait/

[30] See for example WorkSafe Victoria 2021 *The Hierarchy of Controls* www.worksafe.vic.gov.au/hierarchy-control

figure two). Tools that create protective barriers, such as safe searches, are the last line of defence because every instance of individual failure, either from the tool or the user, leaves people exposed to risk.

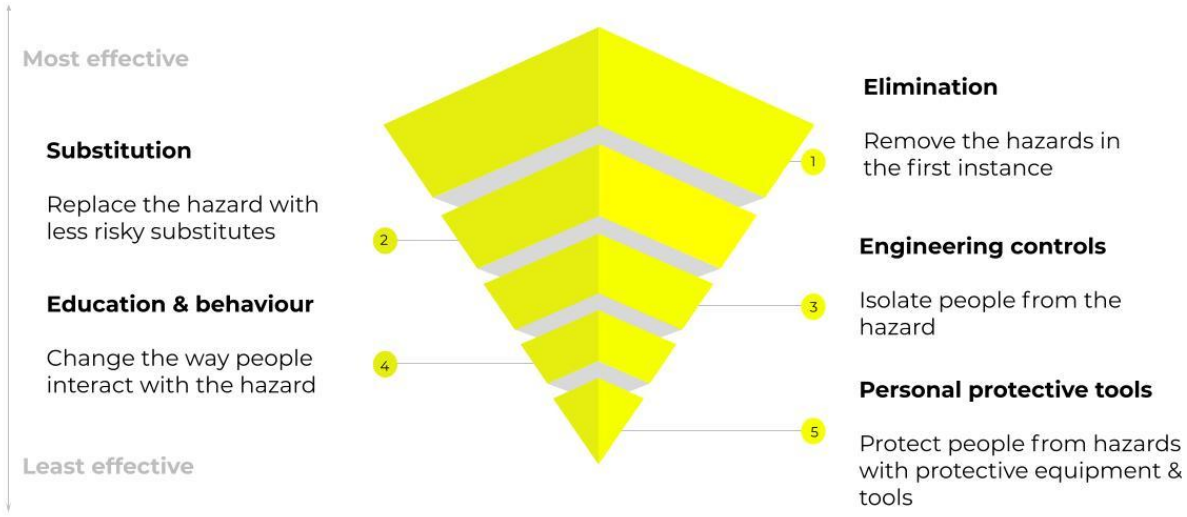

Figure two: The Hierarchy of Hazard Controls

Accountability also requires transparency. Part of the problem of making social media safe is that legislators, regulators, researchers and civil society often do not know enough about the specific mechanics of how platforms work nor their consequences. Requiring transparency through, for example, algorithmic audits and impact statements could help remedy this.

REGULATING FOR TRANSPARENCY: EXAMPLES FROM OTHER JURISDICTIONS

There is an emerging consensus that regulation needs to require transparency from social media platforms, which has emerged in a number of proposed regulations. For example:

- UK: The Online Safety Bill[31] proposes Increased information gathering powers for regulators and investigations powers including power to interview and enter and inspection, and annual transparency reports
- US: The proposal for a Platform Accountability and Consumer Transparency Act[32] including compelling social media platforms to share data with researchers and establish a regulatory Commission on transparency

[31] *Draft Online Safety Bill* 2021, UK https://assets.publishing.service.gov.uk/government/uploads/system/
[32] *Platform Accountability and Consumer Transparency Act 2021, USA* www.congress.gov/bill/116th-congress/senate-bill/4066/all-info

- EU: The Digital Service Act [33] proposes that regulators and vetted academic researchers must be able to access data from large platforms, and that platforms must produce six-monthly, public transparency reports

REGULATING FOR ACCOUNTABILITY: EXAMPLES FROM OTHER JURISDICTIONS

International regulation is increasingly driving towards holding tech companies responsible for any risks that create, and accountable for any harms that occur. For example, the UK's Online Safety Bill uses the principle of a 'duty of care' to create accountability.

The proposal for a Duty of Care in tech regulation was extensively developed Professor Lorna Woods and Will Perrin at the Carnegie Trust, and has support from Australian academics such as Katharine Gelber at QUT[34]. Broadly speaking, the proposals suggest that service providers should be held responsible for their online spaces in the same way that property owners are responsible for physical spaces, and that service providers should have a duty of care to those using their platforms. Professor Woods' has argued that a statutory duty of care would be 'simple, broadly based and largely future-proof', much like long-enduring occupational health and safety regulations which have adopted this approach[35].

The emerging Online Safety Bill in the UK places multiple duties of care on regulated services to reduce the risks in their content and operations[36]. These include risk assessment and mitigation processes, as well as transparency and accountability requirements.. Combined, these duties oblige:

> *'service providers to do particular things, such as undertake risk assessments, to comply with safety duties in respect of illegal content, content that is harmful to children and content that is harmful to adults and other duties, for example in respect of journalistic content. … They are things that providers are required to do to satisfy the regulator. They are not duties to people who use their platforms, and they are not designed to create new grounds for individuals to take providers to court[37].'*

---

[33] Recital 60, European Commission 2020 *Proposal For A Regulation of the European Parliament & of the Council on a Single Market For Digital Services (Digital Services Act) & Amending Directive* 2000/31/ https://oeil.secure.europarl.europa.eu/oeil/popups/ficheprocedure.do?lang=en&reference=2020/0361

[34] Katharine Gelber 2021 'A better way to regulate online hate speech: require social media companies to bear a duty of care to users' *The Conversation* https://theconversation.com/a-better-way-to-regulate-online-hate-speech-require-social-media-companies-to-bear-a-duty-of-care-to-users-163808

[35] Lorna Woods & William Perrin UK 2019 *Online harm reduction – a statutory duty of care and regulator* Carnegie Trust *https://www.carnegieuktrust.org.uk/publications/online-harm-reduction-a-statutory-duty-of-care*

[36] *Draft Online Safety Bill* 2021, UK, link

[37] Joint Committee on the Draft Online Safety Bill 2021 *Draft Online Safety Bill: Report of Sessions 2021-22* https://committees.parliament.uk/publications/8206/documents/84092/default/

4. Ensure the regulatory framework is comprehensive

The rapid growth of the technology has seen Australia's issue-by-issue (e.g. 'cyber bullying', 'image-based abuse' etc), sector-by-sector (e.g. 'social media platforms' 'messaging services' etc) regulatory framework struggle to keep pace. Many new and emergent technologies are missed, and innovative companies straddling the gaps between existing industry definitions are inappropriately regulated.

4.1 GAPS BETWEEN INDUSTRIES AND SERVICES

Australian regulation often takes a sector-by-sector approach, which can fail to adequately address the shared functionalities and integration between the social media sector and multiple other industries. The most obvious of these issues is the integration of traditional media and social media platforms, but equally complicated functionalities exist between social media platforms and data brokers, other online services, the advertising sector, the broader telecommunications industry, and increasingly emergency services and health and social care services as they become central to public messaging campaigns (among others).

Current legislation oftens fails to reflect these integrations and diverse functionalities. Using Roblox, an online kids game, as an example highlights the sorts of peculiarities this can lead to. The current definition of a 'social media' company (as laid out in the *Online Safety Act* and the proposed Enhancing Online Privacy Bill) would fail to cover Roblox. Roblox allows the creation of personal avatars; facilitates and encourages interaction and communication between users, and; allows users to create and share games for others to play. Because users do not post content *per se* — they 'post' games — they are unlikely to be considered a social media platform under the existing definition. Roblox is however covered under the *Online Safety Act* as a 'relevant electronic service' as it facilitates messaging and game play between users. But it would not be covered by the proposed Enhancing Online Privacy Act, unless 2.5 million Australians logged on making it a 'large online platform' under the Bill. Roblox is a platform used mostly by kids, only 30% of their global audience is over 16[38]. There are a total of 3.6m 5-16 year olds in Australia[39], making the 2.5m threshold for coverage improbable. What does this mean? Because Roblox allows users to 'create and share games' rather than 'create and share content', kids may be protected against cyberbullying with regulation, but may not be protected from exploitative data practises or privacy incursions unless almost all Australia's younger kids join the platform, in which case they then may be protected with regulation again. This is regulatory bingo.

[38] Jessica Clement 2021 *Distribution of Roblox games users worldwide as of September 2020, by age* www.statista.com/statistics/1190869/roblox-games-users-global-distribution-age/
[39] Australian Bureau of Statistics 2021 *National, state and territory population, Jun 2021* www.abs.gov.au/statistics/people/population/national-state-and-territory-population/latest-release#data-downloads-data-cubes

Scope limitations can also create regulatory discrepancies between the digital and non-digital world. For example, the ACCC's Digital Platform Inquiry explored the patchwork of regulations that applied to digital publishers compared to telecoms, radio comms and broadcast industries[40]. The ACCC found that despite providing comparable services, digital platforms often feel outside the scope of existing regulation, such as press statements of principles or legislation around advertising gambling and medicine.

Likewise, some exemptions in Australian regulation have not kept pace with the changing digital world. The use of turnover thresholds, such as exempting business above or below $3m or $10m annual turnover, are blunt and do not reflect the emerging nature of the tech industry. In particular, small tech startups can create significant risks for Australians but can often be overlooked. For example, the *Privacy Act* places obligations on businesses with an annual turnover of over $3m but exempts those under this threshold — even those handling significant amounts of data. Blunt exemptions often miss risks, and can leave large companies with safe practice facing a high regulatory burden, while extremely risky small companies continue to deliver harmful products and services.

4.2 GAPS IN EMERGENT TECHNOLOGIES

Likewise, the issue-by-issue, sector-by-sector approach cannot anticipate risks created by innovations and emergent technologies. This has left recent innovations unregulated in Australia, including for example; surveillance advertising and ad delivery systems; AI; blockchain and its integration across systems, and neural technologies.

These gaps suggest that the current approach is unable to future-proof the regulatory framework, and that as technologies evolve, more and more gaps will emerge. Risk focused, systemic models may be more successful at future proofing themselves.

Australians use a wide range of digital services, and seamlessly move between technologies, sectors and companies of all sizes. Their safety should be ensured across their whole digital ecosystem. Gaps and exclusions within Australian regulation have often left Australians reliant on foreign legislation for protection.

[40] Table 4.1, ACCC 2019 *Digital Platforms Inquiry: Final Report*
www.accc.gov.au/system/files/Digital%20Platforms%20Inquiry%20-%20Final%20report%20-%20part%201.pdf

5. Ensure regulation is strong and enforced

Big tech poses big risks and necessitates a robust regulatory response. However, because Australia has to date engaged self- and co-regulatory models by default, our regulatory framework has often failed to reduce risks as rigorously as they otherwise may have.

Future regulation needs to start from the premise that self- and co-regulation will not be sufficient for the social media sector. Reset Australia believes self- and co-regulation have a role to play in the Australian regulatory landscape at large, but that unfortunately the risks posed by the digital environment are:

- High impact, and include significant public health and community safety concerns
- Significant to the community, and the public has an appetite for the certainty of robust regulations
- Unable to be adequately dealt with by lighter touch regulations. The social media sector has demonstrated a track record of systemic compliance issues, including multiple breaches of existing legislation and a generally anaemic response to self-regulation

This warrants a pivot towards primary and subordinate legislation and regulation for the sector. Appendix 3 documents our rationale for recommending an end- to self and co-regulation for the tech sector in more detail.

Alongside strengthening existing regulation, regulators need to be resourced and enabled to enforce this. This includes the ability to fully utilise existing regulation as well as any new legislation proposed.

STRONG & ENFORCED REGULATION: EXAMPLES FROM OTHER JURISDICTIONS

- International developments indicate a shift away from self- and co-regulatory mechanisms towards ‘black letter law’. For example, the DSA upgrades obligations from the voluntary Disinformation Code 2018 into binding legislation.

- Many are further empowering and enabling existing regulators:
  - In the EU, the Digital Services Act proposes new enforcement powers including the ability to order the cessation of infringements, levy fines of up to 6% of global annual turnover as well as periodic penalty payments of up to 5% of average global daily turnover, and accept binding commitments
  - In the UK, the draft Online Safety Bill proposes enforcement powers including directions for improvement, notices of non-compliance, and fiscal penalties like civil fines up to £18 million or 10% of worldwide revenue, and business disruption measures

- Some jurisdictions are establishing new regulators or regulatory functions. For example; Canada has proposed establishing both a new Digital Safety Commissioner and Digital Recourse Council (to handle complaints); Ireland is looking to establish an Online Safety Commission, as part of a broader Media Commission; a number of proposed regulations in the US suggest adding new divisions to the FTC, such as a Youth Privacy and Marketing Division as part of proposals to update COPPA[41], and; in the UK the Online Safety Act will hand over new powers to Ofcom. Just as gaps in regulations themselves need to be addressed, so to do gaps between regulators.

- The ability of regulators to enforce requirements depends on some extent to the level of resourcing they have available. Some Australian regulators are not funded to the same extent as their international counterparts.

Approximate funding per person, in AUD, of different Information Commissioners	
$1.11pp	Office of the Australian Information Commissioner. Australia Based on an annual budget $28,487,000 for 2021-22, Australian population of 25,739,256 in 2021
$1.96pp	Information Commissioner's Office, UK Based on an annual budget £70,625,526 for 2021-22, UK population of 67,081,000 in 2020
$6.04pp	Data Protection Commission, Ireland Based on an annual budget €19,128,000 for 2021-22, Irish population of 5,011,500 in 2021 (Ireland also has EU wide data protection functions)

Reconsidering the powers and resourcing of regulators needs to be part of any attempt to ensure Australia's regulatory frameworks can adequately tackle the risks posed by Big Tech.

[41] Proposed *Children and Teens Online Privacy Protection Act* 2021, US
https://www.congress.gov/bill/117th-congress/senate-bill/1628/text?r=2&s=2

Recommendations

Pivoting to a systemic focus on tech regulation would better serve Australians. In the short and medium term, any proposed legislation or regulation can be evaluated to see how aligned it is with this longer term pivot. This includes asking:

- Does the policy proposal focus on eliminating risks from systems and processes? Does it create duty/duties of care? Does it require risk assessments and mitigations?
- Does the proposal adequately address societal and community harms? Does it address the specific risks posed beyond risks to the individual? Will it address disinformation and hate speech, for example?
- Does the proposal increase transparency and accountability? Does it include responsibilities on platforms that enable public and regulatory oversight?
- Does the proposal create comprehensive coverage? Does it apply to the broadest range of digital services or will it create arbitrary boundaries in protections?
- Does the proposal include adequate regulatory oversight, and is the regulator empowered and resourced for this responsibility?

In the longer term, we could move towards realising this policy landscape through a series of steps, including:

- Building out and expanding our *Online Safety Act* over time to include:
    - A more systemic focus on duties to reduce risks in systems and processes (right across the service, and including for example algorithms and ad delivery systems);
    - Expanding the definition of risks to include community and societal risks, which necessitates an enhanced focus on mis/disinformation and hate speech
    - Requiring enhanced duties of care for accountability, and including requirements for transparency measures
    - Replacing voluntary and co-regulatory codes with upstream obligations in the Act
    - Ensuring that the broadest range of digital services remains covered, with risk based additional obligations. (The scope of the *Online Safety Act* is already very broad, and this provides a potential model for other regulations)
- Expanding our *Privacy Act* and Enhancing Online Privacy Act to:
    - Address a broader definition of personal data to cover metadata and other new forms of data fuelling the new digital world
    - Adopt a systemic focus on reducing the risks created through the processing of data
    - Apply to the broadest range of digital service providers with risk based additional obligations
    - Replace voluntary and co-regulatory codes with upstream obligations in the Act
- Better joining up and resourcing of our regulators. Potentially, a joint digital regulation unit or ‘one stop ombud’ based in the ACCC to oversee data and safety alongside consumer affairs.

This would create a more streamlined approach to regulation, replacing multiple disjointed obligations with more aligned upstream duties and reducing the regulatory burden on Australia's successful tech industry. It would also be interoperable with emerging international requirements, ensuring Australian industry could expand into international markets with minimal regulatory friction.

Appendices

1. WHAT A SYSTEMIC, RISK BASED APPROACH LOOKS LIKE FOR CHILDREN

Exploring what a systemic, risk-based approach looks like for children and young people highlights the range of contextual risks that are inadequately addressed within existing frameworks. The child online safety sector has a commonly used typology that characterises the range of online harms children face; the 4Cs[42]. Figure one contrasts the 4Cs with our regulatory framework.

RISK	SOME OF THE CURRENT REGULATORY FRAMEWORK	GAPS IN FRAMEWORK
Content — risk of exposure to inappropriate content. For example, risks of exposure to violent content, racist content, pornography, sexualised imagery and mis & disinformation	The *Online Safety Act 2021* is establishing frameworks and Codes around class 1 and 2 materials, as well as developing a Restrictive Access System to limit access to age inappropriate materials like pornography. Violent online material may be addressed by the *Sharing Abhorrent Violent Material Act* 2019	Regulation focuses on individual pieces of content, and overlooks the role of platforms in promoting harmful content to children (via algorithms, for example. Hate speech, mis & disinformation are not adequately addressed in the current framework, but can be harmful
Contact — risks of making inappropriate contact with others. E.g. Risks of exposure to online grooming, stalking & extremist recruitment	A number of online laws exist that address contact risks, from the *Criminal Code Amendment (Protecting Minors Online) Act* 2017 to laws around terrorist recruitment. Some of the *Online Safety Act's* co-regulatory codes around ensuring user safety may address ways platforms can reduce contact risks. These are as yet unpublished and will be authored by industry	Existing legislation remedies some harms but does not mitigate risks. While they may criminalise individuals who make inappropriate contact, they do not require platforms to stop recommending adult strangers as 'friends' or 'followers' or prevent platforms enabling adult accounts to message children's accounts for example
Contract / Commercial — risks arising from inappropriate commercial activities and contract exploitation. E.g. risks of identity theft, gambling, profiling bias, surveillance advertising, persuasive design	Children's data is protected as adult's data under the *Privacy Act 1988*, which may reduce the risk of identity fraud. The Online Privacy Code may reduce commercial risks to children's data, but it is yet to be published and will most likely be authored by industry. The Restrictive Access System may restrict gambling (but may miss loot boxes in games).	The use of children's data poses significant risks, and it is unlikely that an industry drafted code — penned by a sector that funds itself through the commercial exploitation of data — will draft a code that puts children's best interests first. There is no regulation in Australia that addresses persuasive design
Conduct — risks associated with inappropriate behaviour. E.g. bullying, trolling, joining harmful groups (e.g anti-vax)	The *Online Safety Act* includes specific provisions around cyber-bullying for children under 18. This includes taking down content that is deemed cyber bullying, and where the perpetrator is a child, the regulator is able to require apologies	Engagement with harmful communities falls outside the scope of current regulatory frameworks

[42] Sonia Livingstone & Mariya Stoilova 2021 *The 4Cs: Classifying Online Risk to Children, CO:RE Short Report Series on Key Topics* doi.org/10.21241/ssoar.71817

2. AUSTRALIAN EXAMPLES OF DISINFORMATION

OVERVIEW	DATE	DESCRIPTION
Thousands of American Bots on Twitter engaged in Australian electoral discussions[43]	2019	A QUT study examined 54,000 accounts of around 130,000 Twitter users active during and after the 2019 Australian Federal Election and found that 13% of accounts were 'very likely' to be bots, with the majority originating from New York. This is estimated to be more than double the rate of bot accounts in the US presidential election. This was done through an AI program Botometer - which looks for signs such as tweeting frequently 24 hours a day, tweeting at regular intervals, usernames with lots of numbers and whether their followers also appeared to be bots. New accounts created during the election campaign were more likely to be bots.
Anti-Labour disinformation and fake news on WeChat[44]	May 2019	The influential Chinese social media site WeChat was presenting challenges to Labor as a wave of fake news posts and doctored accounts targeted the Shorten campaign on issues such as Safe Schools, taxes and refugee policy. Many of the posts were unauthorised, so it was difficult to know who was responsible for them. However: • A doctored tweet was found on multiple WeChat groups, posted by a Liberal Party member. The tweet was doctored to look like Bill Shorten had said "Immigration of people from the Middle East is the future Australia needs" • Another WeChat account was traced back to former Liberal MP, whose former state seat had high numbers of Chinese-Australian voters. The account was registered in the Sept 2017 as "MichaelGidleyMP" but changed in 2019 to "Victoria Brief Talk" then again to "Australia Brief Talk". The account posted multiple instances of disinformation, such as claiming that under Labour retirees would pay additional tax, or face extra taxes on house sales of $30,000.
Anti-Liberals propaganda from social media accounts affiliated with the Chinese Communist Party[45]	May 2019	There is evidence of anti-liberal propaganda which has the potential to be chinese state interference. Across a period of five months from November 2018 to March 2019, researchers analysed the Australian content on 47 of the most visited WeChat Official accounts in Mainland China, 29 of which were aligned with the CCP. Researchers found that there was a clear "anti-Liberal story" coming from social media accounts, many of which have close affiliations to the Chinese Government. The posts also criticise Australia's involvement in the Five Eyes alliance.
Billions of Facebook accounts publishing disinformation active during Australian	May 2019	In their response to a Joint Committee on Electoral Matters Facebook outlines that it removed 2.2bn fake accounts between January and March 2019 engaging in coordinated inauthentic behaviour. Facebook stated that they*" do not believe that it's an appropriate role for us to be the arbiter of truth over content shared by ordinary Australians or to referee political debates and prevent a politician's speech from*

[43] Felicity Caldwell 2019 'Bots stormed Twitter in their thousands 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

[44] Yan Zhuang & Farrah Tomazin 2019 'Labor asks questions of WeChat over doctored accounts, 'fake news' *Sydney Morning Herald* https://www.smh.com.au/national/labor-asks-questions-of-wechat-over-doctored-accounts-fake-news-20190506-p51kkj.html

[45] Steve Cannane 2019 Chinese media mocks Australia and Prime Minister in WeChat posts *ABC News* https://www.google.com/url?q=https://www.abc.net.au/news/2019-05-09/pm-targeted-by-chinese-communist-party-related-wechat-accounts/11092238&sa=D&source=docs&ust=1641786278156843&usg=AOvVaw1mrFcD73xejHL-oxGD5dpC

election[46]		*reaching its audience and being subject to public debate and scrutiny."*
Kosovan scammers target 'coordinated inauthentic behaviour' at Australians on a massive scale, some of which is political[47]	2019	A network of Facebook pages run out of the Balkans profited from the manipulation of Australian public sentiment, and ran over an election period. Posts were designed to provoke outrage on hot button issues, driving clicks to stolen articles in order to earn revenue from Facebook's ad network. The scale of the operations were immense, the pages had a combined fanbase of 130,000-plus, built up over several years. The oldest and most popular page, "Australians against Sharia", had been publishing since June 2013, had over 67,000 fans and reposted memes attacking Labor politicians Bill Shorten, Penny Wong and Julia Gillard, the Greens' Sarah Hanson-Young and the Liberal Party's Julie Bishop. Facebook removed these pages for violating their policies by engaging in "coordinated inauthentic behaviour".
Russia's Internet Research Agency targeting Australian politics between 2015 and 2017 on Twitter[48]	2015 - 2017	Twitter identified 3,841 accounts suspected of operating out of the Internet Research Agency in St Petersburg. A number of these accounts targeted Australian politics. Researchers from Clemson University in the US analysed 3 million tweets, finding many accounts targeting Australian politics, particularly the Australian response to the downing of flight MH17. Some 5,000 tweets mention the terms "#auspol", "Australia" or "MH17" – with "Australia" the most common of the three. • An initial spike of activity focusing on MH17 correlated with the Australian deployment of fighter aircraft to operate in Syrian airspace, where Russian aircraft were operational. • A second spike in Feb 2017 centred around a hashtag game, where Russian accounts encouraged people to come up with Australian names for popular US television programs. While this may seem like innocent fun, it is also an old spy craft technique to recruit "assets" on neutral, non-political terms before escalating to political topics.
Bushfire disinformation is spread on Twitter[49]	2020	Hashtag #ArsonEmergency became the focal point of the bushfire crisis in the beginning of 2020, and pushed a narrative that the cause of the fires was arson. QUT social media analyst Timothy Graham studied 300 twitter accounts looking for inauthentic behaviour driving the #ArsonEmergency hashtag, and many were found to be behaving 'suspiciously'.
Covid misinformation is spread widely across social media[50]	2020	Communications Minister Paul Fletcher warned Australians to be sceptical of what they read online as Covid disinformation spread. Posts suggested it was deliberately released as directed people not to consume certain food or visit particular areas in Australia.

[46] Katharine Murpy 2019 ' Facebook removed 'coordinated inauthentic behaviour' during Australian election' *The Guardian* https://www.theguardian.com/australia-news/2019/oct/23/facebook-removed-coordinated-inauthentic-behaviour-during-australian-election
[47] Michael Workman & Stephen Hutcheon 2019 'Facebook trolls and scammers from Kosovo are manipulating Australian users' *ABC News* https://www.abc.net.au/news/2019-03-15/trolls-from-kosovo-are-manipulating-australian-facebook-pages/10892680
[48] Tom Sear & Micael Jensen 2018 'Russian trolls targeted Australian voters on Twitter via #auspol and #MH17' *The Conversation* https://theconversation.com/russian-trolls-targeted-australian-voters-on-twitter-via-auspol-and-mh17-101386
[49] Timothy Graham & Tobian Keller 2020 'Bushfires, bots and arson claims: Australia flung in the global disinformation spotlight' *The Conversation* https://theconversation.com/bushfires-bots-and-arson-claims-australia-flung-in-the-global-disinformation-spotlight-129556
[50] Zoe Samios and Dana McCauley 2020 Minister urges scepticism as fake virus news spreads' Sydney Morning Herald https://www.smh.com.au/business/companies/minister-urges-scepticism-as-fake-virus-news-spreads-20200128-p53vjn.html

3. ASSESSMENT OF RISK WARRANTING ESCALATION FROM SELF- AND CO-REGULATION TO PRIMARY AND SUBORDINATE LEGISLATION

Australia has developed a multi-path approach to industry regulation, using self-regulation, quasi-regulation, co-regulation and 'black letter law' (or explicit regulation by primary and subordinate legislation). Since the mid '90s, this multi-path approach has facilitated a range of regulatory responses, some 'light' and some 'hard touch' to different industries and issues.

While over a decade old, the *Best Practice Regulation Handbook*[51] outlined considerations to assess which path is appropriate for each industry/issue, that are still helpful prompts for reflection. The handbook suggests an evaluation of the options should consider:

- The level of risks and significance posed by the potential concern, noting that major public health and safety issues warrant explicit government regulation
- The community appetite for the certainty of legal sanctions, noting that self regulation is only feasible where there is no particular community interest
- The ability of the market to address the concern, noting that where there is "a systemic compliance problem with a history of intractable disputes and repeated or flagrant breaches of fair trading principles, and no possibility of effective sanctions being applied" explicit regulation is required.

We believe that social media has exceeded any reasonable threshold for explicit government regulation across all three considerations.

1. **The level of risks posed by social media platforms:** Social media platforms can create significant risks, including major public health risks. Taking Facebook and the pandemic as an example, Australia witnessed the enabling and promotion of harmful content and discussions. Both membership numbers and engagement among groups peddling 'anti-vaxx' and vaccine hesitant content grew across the pandemic in Australia[52]. We also saw the rise and promotion of 'anti vaxx' influencers, with more than 100 Instagram accounts promoting anti-vaxx content to more than 6 million users[53]. There were ,and will continue to be, deadly consequences while these risks continue.

2. **The community wants and expects the certainty provided by regulation:** There are now legitimate community expectations of explicit regulation of Big Tech in Australia. Last year, a Lowy Institute poll found that 90% of Australians think that the influence social media companies have is an important or critical threat to the vital interests of Australia[54]. And indeed, a poll by the Australian

[51] Australian Government 2010 *Best Practice Regulation Handbook* Canberra
[52] Reset Tech Australia 2021 *Anti-vaccination & vaccine hesitant narratives intensify in Australian Facebook Groups* https://au.reset.tech/uploads/resetaustralia_social-listening_report_100521-1.pdf
[53] Jasper Jackson & Alexandra Heal 2020 'Instagraft: Covid conspiracy theorists selling silver spray and $50 seawater' *Bureau of Investigative Journalism* www.thebureauinvestigates.com/stories/2021-04-11/instagraft-covid-conspiracy-theorists-selling-silver-spray-and-50-seawater?mc_cid=faeeac9b83&mc_eid=ae64430abe
[54] Lowry Institute 2021 *Lowry Institute Poll* https://poll.lowyinstitute.org/charts/threats-australias-vital-interests/

Financial Review in late 2020 found that 77% of Australians felt that BigTech should face stronger Government regulations[55]. The scale and depth of the public's concerns warrants the strongest possible regulatory response.

3. **The social media sector has demonstrated systemic compliance problems:** While many sectors have worked hard to deserve the benefit of 'light touch' regulations, the social media sector has demonstrably not. For example:

    - YouTube settled a case for $170m USD with the FTC in 2019 for using children's data without necessary parental consent[56] and are currently facing a £2b 'class action' for unlawfully tracking and collecting children's data[57]. Google, their parent company, has also been fined for multiple breaches of existing regulation, including a €500m fine for acting in bad faith around EU copyright directives in France[58], €7m for failing to meet requirements around GDPR in Sweden[59] and €220m for anti competitive practices in their advertising systems in France again[60]. Earlier this year, the Texas Attorney General accused Google of deliberately stalling efforts to strengthen children's online privacy laws in the US, and documented Google executives 'bragging' about stalling EU attempts at improving consumer privacy[61].
    - Facebook has faced many fines, including a $5b USD penalty from the FTC for breaching consumer privacy regulations[62] and a $5m USD to settle civil rights lawsuits claiming the company's advertising system excluded people from seeing housing ads based on age, gender and race[63].
    - TikTok has also had its fair share of fines, settling a case for $5.7 m USD with the FTC in 2019 for using children's data without the necessary parental consent[64], and were fined €750k in the Netherlands over GDPR compliance[65]. They are currently facing a £1b plus lawsuit led by the UK's former Children's Commissioner for excessive data collection practices[66].

---

[55] Paul Smith 2020 'Big Tech on the Nose' *Australian Financial Review* www.afr.com/technology/big-tech-on-the-nose-as-aussies-demand-accountability-and-tougher-laws-20201030-p56a93

[56] FTC 2019 'Google and YouTube will Pay Record $170m for Alleged Violations of Children's Privacy Law' www.ftc.gov/news-events/press-releases/2019/09/google-youtube-will-pay-record-170-million-alleged-violations

[57] YouTube Data Claim 2020 'YouTube Data Claim' www.youtubedataclaim.co.uk/

[58] Ian Carlos Campbell 2021 'Google fined €500 million in France over bad faith negotiations with news outlets' *The Verge* www.theverge.com/2021/7/13/22575647/google-fine-500-million-french-authorities-news-showcase

[59] Vincent Manancourt 2020 'Google to appeal Swedish data watchdog' *Politico* www.politico.com/news/2020/03/11/google-to-appeal-swedish-data-watchdog-7m-fine-125460

[60] Simon Read 2021 'Google Fined €220m in France' *BBC* https://www.bbc.com/news/business-57383867

[61] Leah Nylen 2021 'Google sought feelow tech giants help is stalling kids privacy protections' *Politico* www.politico.com/news/2021/10/22/google-kids-privacy-protections-tech-giants-516834

[62] FTC 2019 *FTC imposes $5 Billion Penalty* www.ftc.gov/news-events/press-releases/2019/07/ftc-imposes-5-billion-penalty-sweeping-new-privacy

[63] Brakkton Booker 2019 'After Lawsuite, Facebook Announces Changes' *NPR* www.npr.org/2019/03/19/704831866/after-lawsuits-facebook-announces-changes-to-alleged-discriminatory-ad-targeting

[64] FTC 2019 *Video Social Networking App Settles* www.ftc.gov/news-events/press-releases/2019/02/video-social-networking-app-musically-agrees-settle-ft

[65] Dutch News 2021'Dutch Privacy Watchdog Fines TikTok' *Dutch News* www.dutchnews.nl/news/2021/07/dutch-privacy-watchdog-fines-tiktok-e750000-after-privacy-probe

[66] BBC 2021 'TikTok sued for billions over use of children's data' *BBC* www.bbc.co.uk/news/technology-56815480

Beyond compliance with existing regulations, at times the sector appears to actively resist 'doing the right thing'. For example, back in 2016, the Wall Street Journal found an internal Facebook presentation documenting that they know their platform was hosting a large number of extremist groups and promoting them to its users: "64% of all extremist group joins are due to our recommendation tools," the presentation said[67]. It was only in the wake of the insurrection in January 2021 that Mark Zuckerberg announced that the company will no longer recommend civic and political groups to its users.

This does not reflect a series of unrelated incidents. Most of these companies are publicly listed entities obligated to act in shareholder's best interests. Without legal requirements insisting that they prioritise user safety, they are bound to continue to prioritise shareholder profits.

[67] Jeff Horwitz & Deepa Seetharaman 2020 'Facebook Executives Shut Down Efforts to Make the Site Less Divisive' *Wall Street Journal*
www.wsj.com/articles/facebook-knows-it-encourages-division-top-executives-nixed-solutions-11590507