ZOE
DANIEL
MP
Independent Federal Member for Goldstein

Submission: Statutory Review of the Online Safety Act 2021

Australia is often regarded as a world leader in online safety regulation. Our existing content removal regulatory model, headed by the *eSafety Commissioner,* enjoys bipartisan support in the parliament and consensus in our national discourse. My submission to the Statutory Review of the Online Safety Act 2021 argues that this view is dated, on the basis that our regulatory regime has been superseded by more modern approaches adopted by comparable OECD jurisdictions. When compared to these jurisdictions, Australian regulation in fact lags behind.

If Australia is to modernise its approach to digital platform regulation and meaningfully curtail the social risks propagated and exacerbated by digital platforms, government must urgently turn its attention to implementing a version of the emerging 'systems'-based approach as has been legislated in the European Union (EU) and United Kingdom (UK). The systems regulatory model focuses on the systems (i.e., 'algorithms' and 'recommender systems') used by digital platforms in the administration and management of their services at scale. Without such reform, the Australian Government risks allowing the broad range of risks digital platforms present to our children, adults, and society to continue. Reform of this nature could be achieved via amendment to existing legislation, primarily the Online Safety Act 2021 (OSA).

Relationship to Eating Disorders and Parliamentary Working Group

In September 2023, I convened a working group to analyse the relationship between social media and eating disorder prevalence in Australia, and to identify policy options for reform. The working group was comprised of leading Australian experts in the relevant policy fields, in addition to representatives from *Meta*. On 28 May 2023, I hosted a follow-up roundtable at Parliament House, during which the working group released its *Body Image and Social Media* report (the Report).

Eating Disorders provide just one example of harm that can be linked to social media use, but they provide a good starting point for policy development and mitigation strategies that can be applied to other forms of harm.

The Report identified social media use as a "major factor escalating the risk" of eating disorders and noted that eating disorders among youth aged 10-19 rose by 85% since 2012, roughly coinciding with the escalation in social media use. The Report's 21 recommendations centred thematically on measures to promptly remove pro-eating disorder content (i.e., extreme diet or weight loss), regulation of the systems digital platforms operate, transparency measures, and non-compliance penalties. Many of these recommendations are simple to implement and would serve to mitigate eating disorder prevalence in advance of broader reform.

Of highlight:

- Modify the *Online Safety (Basic Online Safety Expectations) Determination 2022* to require digital platforms take reasonable steps to remove pro-eating disorder content from their platforms, including advertisements, with penalties for inaction.
- Implement an overarching and enforceable duty of care onto the operation of digital platforms in Australia, as is applied in other areas of law such as Occupational Health and Safety, medicine, hospitality, education, and professional services.
- Risk assessments and risk mitigation obligations on the harms which could be reasonably caused by all of the systems and elements digital platforms operate.
- Expand the means by which Australian users are able to report to the *eSafety Commissioner* content which could negatively impact their body image.
- Require digital platforms provide users the opportunity to easily reset their personal 'algorithms,' including the ability to opt-out of specific or most content recommendation.
- Mandate digital platforms to feature diverse physical appearances (i.e., diverse body sizes, shapes, genders, colours, and abilities) among advertisements approved by an operator.

- Restrict the use of beauty filters to Australians over the age of 18 years; and
- Federal funding for harm minimisation and prevention, including targeted scientific research.

The Report's public release coincided with the publication of multiple Australian scientific research studies which demonstrated a link between addictive social media use and changes in human physiology and behaviour. One study reviewed functional magnetic resonance imaging (fMRI) data to identify increased neural activity in regions of the brain when adolescent participants were at rest and a decrease in functional connectivity in parts of the brain responsible for memory and decision making.[1] Another study identified an increase in social media use by vulnerable patients suffering withdrawal symptoms from antidepressant medications.[2]

Viewing the harms caused by social media through the lens of eating disorders is useful but the social harms propagated via digital platforms are not limited to mental health. Digital platforms are currently incentivised to amplify a broad array of harmful and 'outrage' content, in part by the way their recommender and other systems and features are designed to maximise user attention. This occurs within a sophisticated digital architecture and business model which benefits the longer a user spends time on and shares content across a digital platform.

Whilst I recognise that the emergence and entrenchment of social media has occurred within a broader context of a 21st century transition toward digitalisation, this does not absolve public policymakers of regulatory responsibility. Options are currently available to the Australian Government which would assist in shaping digital platforms into the healthier ecosystems of information sharing and socialisation urgently needed by Australians of all age groups.

Limitations of Current Discourse and Government Response

At time of writing, much of the ongoing political discourse on online safety centres around the question of whether access to social media should be restricted by age. There is obvious popular appeal in this approach, however it is flawed in two ways. First, children and adolescents are likely to either misrepresent their age or quickly identify means to evade an unsophisticated age verification system. Second, it ignores available policy options which would shape digital platforms into safe digital spaces for our children and adolescents to inhabit. A more sophisticated alternative would be to mandate digital platforms to turn off the use of recommender and/or other systems and elements up to a certain age for the platforms of most risk to adolescent wellbeing. This provides adolescents with an introduction to modern digital technology without being profiled and served content potentially harmful to their physiological and emotional development. This measure could be implemented in advance of broader online safety reform akin to the EU and UK.

On 30 May 2024, the Minister for Communications announced a new determination to amend the OSA's Basic Online Safety Expectations (BOSE). Inclusion of the rights of the child in the design and operation of a digital platform, including transparency reporting, represents a symbolic step forward toward a systemic approach to online safety in Australia. These measures, however, remain limited; the power of the BOSE relates mainly to illegal material and behaviour, such as child sexual exploitation or terrorism. An overarching duty of care supported by regular mandatory risk reporting and robust features for user empowerment remains the most comprehensive regulatory pathway for Australia, modelled from the UK Online Safety Act and EU Digital Services Act.

To-date, public discourse and governmental policy responses on child safety age verification exclude the risk digital platforms pose to adults of all age groups. An individual does not arrive at any age and suddenly develop

---

[1] Chang. M. & Lee. R (2024). Functional Connectivity Changes in the Brain of Adolescents with Internet Addition: A Systemic Literature Review of Imaging Studies *PLOS Mental Health* https://doi.org/10.1371/journal.pmen.0000022

[2] Coe, A., Abid. K, & Kaylor-Hughes. C (2024). Social Media Group Support for Antidepressant Deprescribing: A Mixed-Methods Survey of Patient Experiences *Australian Journal of Primary Health, 30.* https://www.publish.csiro.au/py/PY23046

resilience to harmful content, addictive design features, and misaligned systems. Adult wellbeing remains highly susceptible to these elements and should not be omitted from regulatory scope.

A further limitation relates to the inherent inability of the eSafety Commissioner-led regulatory mode to operate at scale. The takedown order regulatory approach is modelled from a 20th century media landscape and was not designed for modern 21st century digital communications and information sharing. User-generated content on a digital platform is decentralised and disseminates across a service at a pace which traditional takedown orders, issued by the eSafety Commissioner, cannot manage effectively. The eSafety Commissioner deserves respect for work delivered thus far, however his limitation brings the underlying content-first strategy that guides Australian online safety regulation into question; take down orders still play a role amidst a broader systems-first framework.

The Systems Approach

The following outlines some of the measures which are available to policymakers in the development of an Australian version of a systems approach to online safety. It is not intended to be exhaustive.

An Overarching Duty of Care

At the centre of the systems model of online safety lies the application of an overarching duty of care, which represents a far more comprehensive regulatory approach than its tiered alternative or Australia's current trajectory of iteratively amending the BOSE under existing law. A duty of care regime would significantly expand the purview of Australia's OSA beyond its presently limited focus mainly on material that is abhorrent or illegal. As is intuitively understood by the Australian public, a broad array of social and societal risk is propagated via digital platforms, much of which may be enigmatic and not yet perceptible. The duty of care model would empower Australian regulators to intervene in this broader landscape of risk, and could apply obligations on digital platforms to mitigate foreseeable risk to:

- Illegal material, such as Class 1A & 1B under the existing OSA.
- Cyber bullying of children and abuse of Australian adults (to a greater degree than in the OSA).
- Australian electoral processes and public security.
- Human and civil rights.
- Hate speech and social cohesion.
- Gender-based violence and division.
- Mental health and eating disorders.
- Attention spans and internet addiction.
- Demonstrably false misinformation or disinformation, including foreign interference; and
- Additional risks to Australian society as determinable by the Minister.

Risk Reporting and Risk Mitigation Duties

In the interest of establishing a duty of care regime which is supported by enforceability, digital platforms must also be required to regularly assess the risk that their systems could foreseeably cause to the specified obligations under a duty of care. Platforms should subsequently be required to identify and implement best practice risk mitigation measures in response to each of their self-identified risks. Both assessments must be submitted to regulators for review.

The United Kingdom's Online Safety Act, in tandem with risk reporting and mitigation duties, requires digital platforms to subject themselves to regulatory compliance measures. This includes mandatory compliance with requests for information and to investigate or audit a digital platform.

Extending a digital platform's risk reporting and mitigation duties to encompass "all of their systems and elements" is critical, in recognition of the common misconception that recommender systems alone are the exclusive source of risk on a digital platform. Harm to individuals and society can be propagated via systems and elements of various types: advertisement approval systems; automated user suspension systems; automated content moderation systems; addictive design features; among many others. Whilst many of these systems are indeed necessary for a platform to operate at scale, their actual effectiveness and design ethic remains controversial and laden with risk. As such, regulatory scrutiny is warranted.

User Empowerment Features

On August 25, 2023, the EU's Digital Services Act came into effect, and with it the capability for users of the largest digital platforms to opt-out of personalised content sorting and recommendation. In effect, this empowers European users to receive content populated almost exclusively by content produced in their immediate 'friends' or 'following' network. Whilst this may not be the preference for every Australian user, availability of the option alone would be a tangible forward step for individual choice.

A version of this healthy design feature should similarly be required to be implemented for Australian users. Alternatives to this feature may involve the capability to selectively opt-out of content recommendation for specific profiles of material based on individual preference. These features are reasonably simple for a digital platform's internal application development operations team to engineer for a specific geographical region.

Transparency and Empowerment of Independent Australian Researchers

In the interest of establishing a data-driven regulatory regime and fostering an informed and educated Australian population, the risk reporting and mitigation duties should be made publicly transparent. These reporting documents may be made available after a specified period following their submission to regulators, such as a duration of twelve months.

A key limitation in international online safety discourse and policy is lack of scientific research into social media harm and internet addition. This is, in part, due to the scale involved for studies to deliver conclusive findings. Many of the activities involved for study participants, such as abstention from social media, are often perceived as overly burdensome or unacceptable to adhere to. An Australian version of a systems-first regulatory framework may include openly searchable ad approval and rejection repositories and researcher access to data where it is in the public interest. A further measure may include funding for dedicated scientific research studies, which would also advance international online safety discourse and policy.

Penalties for Non-Compliance

In the context of regulating corporate entities which are among the most profitable in the world, fines are generally insufficient to meaningfully deter or remedy harm. The UK's Online Safety Act resolves this obstacle in two ways and could be adopted by Australia. First, the UK's Office of Communications (Ofcom) is now able to issue substantial penalties of up to £18 million or 10% of global revenue for breaches of the new Online Safety Act. Senior management of digital platform operators may also be liable where non-compliance has been demonstrated. Second, Ofcom may issue the more severe penalty of business disruption measures where a digital platform has demonstrated continued or serious non-compliance. Business disruption involves prohibiting a third party from continuing to provide an ancillary service to a corporate entity regulated under the Online Safety Act. These measures require careful consideration due to the risk of heavy handedness.

Conclusion

In an apparent reaction to the two horrific knife attacks which occurred in Bondi Junction and Wakeley this year, online safety has received heightened attention in Australia's national political discourse. This also prompted the government to bring forward the statutory review of the Online Safety Act 2021. Considering these developments, an opportunity has emerged for the Australian Government to legislate meaningful online safety reform in the interest of our children, adults, and society. Not only is this model of public policy the international best-practice, but it would likely also receive broad political support from the national electorate. I strongly encourage the government to adopt a systems-first regulatory approach as an outcome of this statutory review.

References

Chang. M. & Lee. R (2024). Functional Connectivity Changes in the Brain of Adolescents with Internet Addition: A Systemic Literature Review of Imaging Studies *PLOS Mental Health* *https://doi.org/10.1371/journal.pmen.0000022*

Coe, A., Abid. K, & Kaylor-Hughes. C (2024). Social Media Group Support for Antidepressant Deprescribing: A Mixed-Methods Survey of Patient Experiences *Australian Journal of Primary Health, 30.* *https://www.publish.csiro.au/py/PY23046*

Reset Australia. (2024). *Not Just Algorithms: Assuring User Safety Online with Systemic Regulatory Frameworks*. https://au.reset.tech/uploads/Not-Just-Algorithms-web-230323-V1.0.pdf

Humphry. J. (May 2024) Age Verification for Social Media: Do Kids and Parents Event Want it? *University of Sydney. https://www.sydney.edu.au/news-opinion/news/2024/05/23/age-verification-social-media-do-kids-parents-want-it-expert.html*