

Harm blindness in digital child protection policy: A systematic stakeholder, sentinel and legal analysis of Australia's under-16 social media ban

Travis Gilly,² *The Real Safety AI Foundation, USA*

Rene Mattheus,² *Pattern Interrupt Consulting, South Africa*

Introduction

On 10 December 2025, Australia became the first nation to enforce a blanket prohibition on social media access for users under 16.

The Online Safety Amendment (Social Media Minimum Age) Act 2024 exposes platforms to civil penalties of up to AUD 49.5 million for systemic failures to prevent underage account creation, and has reshaped the regulatory conversation in the UK, EU, Malaysia and South Africa, where the Minister of Communications and Digital Technologies has indicated similar measures are under active consideration.

This paper argues that the Australian ban is not merely an ethically costly policy; it is a legally fragile one. Using the Harm Blindness Framework (HBF) to identify stakeholder damage, the Paediatric Sentinel Effect in Algorithmic Harm to explain why children's experiences predict adult harms, and a doctrinal proportionality analysis grounded in South African constitutional law and the United Nations Convention on the Rights of the Child (UNCRC) and *General comment No. 25*, we demonstrate that the ban fails on its

² The authors contributed equally to this work.

own terms and would fail doctrinal scrutiny in any jurisdiction that applies ordinary child rights, data protection and proportionality tests.

Research questions

1. Does the legislation's framing of children as a monolithic beneficiary class obscure harm to vulnerable subpopulations, including LGBTQ+ youth in unsupportive households and children in abusive homes who rely on digital platforms for identity formation and crisis support?
2. What perverse incentives does the ban create for platforms, parents and young people, and do these undermine the policy's stated child protection objectives?
3. Does the policy address the root cause of harm to children online, namely exploitative platform design optimised for engagement, or does it displace responsibility onto families while leaving harmful business models intact?
4. How does the enforcement and age assurance compliance architecture of the Australian Act stand up against comparative data protection, platform liability and children's rights instruments, and what does that assessment mean for jurisdictions (notably South Africa) considering analogous legislation?

Theoretical framework

The Harm Blindness Framework (HBF) is a systematic stakeholder analysis methodology that identifies preventable harm at four decision points: ideation, design, testing and launch. It defines harm blindness as the systematic failure to identify stakeholder harm, driven by stakeholder myopia, ethical abdication, precedent fallacy and speed pressure.

The Paediatric Sentinel Effect in Algorithmic Harm proposes that children function as sentinel populations for algorithmic systems, experiencing harms earlier and more visibly than adults. Drawn from toxicology and environmental health, the framework rests on three mechanisms: mandatory institutional exposure that reduces children's capacity to opt out; cognitive developmental factors that increase susceptibility to algorithmic manipulation; and a paradox of data protection in which the most legally protected population generates the most behavioural data. Ubuntu philosophy supplies a Global South communal ethic that rejects the individualist framing of digital safety as parental responsibility.

To these we add a doctrinal proportionality lens: Article 3 of the UNCRC and *General comment No. 25* require measures restricting children's digital access to satisfy necessity, proportionality and least restrictive means tests, and not result in children's exclusion from the digital environment. South African jurisprudence places the best

interests of the child as paramount under Section 28(2) of the Constitution, applied contextually to the affected child rather than as an abstract formula (*S v M* (Centre for Child Law as Amicus Curiae) 2008 (3) SA 232 (CC); *De Reuck v Director of Public Prosecutions (Witwatersrand Local Division)* 2004 (1) SA 406 (CC)). Together, these lenses convert a policy critique into a legal assessment against a doctrinal standard.

This analysis sits within the childism literature without being captured by either dominant stream. Young-Bruehl (2012) frames childism as prejudice against children comparable to racism and sexism, legitimating policies that harm children while claiming to protect them. The Australian ban displays this pattern: protective rhetoric masks the exclusion of LGBTQ+ youth, abused children and disabled users from communities they depend on, without consultation. Wall (2008, 2010) frames childism as a critical theory modelled on feminism, in which children's experiences transform ethics and rights rather than being slotted into adult frameworks. The Paediatric Sentinel Effect operationalises this move empirically. Children's algorithmic harms function as leading indicators of risk at the population level. The HBF supplies the diagnostic apparatus Young-Bruehl's account lacks; the Sentinel Effect supplies the predictive engine Wall's framework points toward.

Empirical method

We applied HBF's four checkpoint methodology to the Australian Act using legislative text, parliamentary debate transcripts, e-Safety Commissioner consultation submissions, media coverage and platform policy responses.

For the sentinel analysis, we drew on 40 child-focused policy documents comprising over 7,300 extracted harm terms, cross-referenced with the AI Incidents Database, tracing timeline evidence across biometric surveillance, predictive risk assessment and recommender systems.

For the legal analysis, we assessed the Act against the Commonwealth of Australia's *Age assurance technology trial: Final report* (2025); the UNCRC and *General comment No. 25*; the African Charter on the Rights and Welfare of the Child (OAU Doc. CAB/LEG/24.9/49) (1990); the Constitution of the Republic of South Africa; the Children's Act 38 of 2005 (South Africa); POPIA (Protection of Personal Information Act) No. 4 of 2013 (South Africa); the Information Regulator's 2021 *Guidance note on processing personal information of children*; the Cybercrimes Act No. 19 of 2020; the Films and Publications Amendment Act No. 11 of 2019 (South Africa); and Constitutional Court authority on Section 28(2) and Section 36 limitation analysis.

Findings

The Australian ban exemplifies harm blindness across all four HBF checkpoints. Checkpoint 1 identifies nine stakeholder groups beyond intended beneficiaries, with two at critical severity: LGBTQ+ youth in unsupportive households lose identity-affirming communities, peer support and crisis resources when experiencing suicidal ideation, children in abusive homes lose external lifelines, independent communication channels and access to mandatory reporters, effectively trapping them with their abusers.

Checkpoint 2 identifies four categories of perverse incentive: platforms gain biometric verification data while exploitative design remains untouched; teenagers migrate to less regulated spaces including Discord, Telegram and VPN workarounds; parents develop false security that reduces supervision; and governments claim decisive action without confronting platform business models.

Checkpoint 3 records the failure of meaningful consultation with affected youth, LGBTQ+ advocacy organisations, disability advocates and digital rights experts. Checkpoint 4 projects predictable breaking points including documented harm to vulnerable youth, technical circumvention and international criticism.

The sentinel analysis advances children's rights scholarship beyond welfare framings by treating child-focused harm as a leading indicator of population-level risk. School biometric surveillance in Lockport preceded the first adult wrongful arrest by eight months. The Allegheny Family Screening Tool established administrative data prediction precedents now migrating to adult contexts. Internal Facebook research documented Instagram's impact on adolescent girls before adult platform addiction was clinically recognised.

Under the sentinel framework, the Australian ban is a second-order harm: it does not address algorithmic toxicity in children; rather, it eliminates the sentinel population whose experiences warn of harms heading toward adults.

The legal analysis reveals the doctrinal fragility that the existing welfare critique misses. First, enforcement depends on the Commissioner's ability to impose and collect penalties against foreign domiciled platforms with no local legal presence, a problem also identified in South Africa. In the age verification enforcement context, neither subpoena, enforcement notice nor mutual legal assistance treaty offers a complete structural solution: all face dual criminality, delay and scope limits in a regulatory scheme that is civil or administrative rather than criminal. A statute that cannot be enforced against its primary duty-bearers operates as signalling rather than regulation.

Second, the age assurance compliance paradox is unavoidable: the Commonwealth of Australia's *Age assurance technology trial* confirmed that every available method carries measurable error, and requires either government identity documents or biometric

data from every user. Children's data is therefore processed more intensively, not less, which inverts the protective logic of Section 34 of POPIA and the Information Regulator's 2021 *Guidance note*.

Third, a South African replication would engage Section 36 of the Constitution, requiring justifiable limitation of rights under Sections 16, 18, 28(2) and 9. Fourth, the ban would worsen harms already criminalised under the Cybercrimes Act and Films and Publications Amendment Act, while reducing visibility to mandatory reporters.

Evidence-based alternatives map on to operational South African instruments. Algorithmic transparency attaches to POPIA Section 71, read with Section 5. Addictive design prohibition attaches to Consumer Protection Act Sections 48 and 49 and the Information Regulator's Section 35(2) power. Age-appropriate design standards can be implemented through a Section 60 code of conduct modelled on the UK Children's Code. Platform liability extends Sections 18F to 18H of the Films and Publications Amendment Act. Digital literacy obligations attach to the Children's Act and the Schools Act No. 84 of 1996.

The Act engages UNCRC Article 2 (non-discrimination), Article 12 (participation, given the seven-day parliamentary passage without youth consultation), Articles 13 and 17 (expression and information access), Article 16 (privacy) and Article 19 (protection from violence). Articles 4 and 10 of the African Charter are directly engaged. *General comment No. 25* is explicit that protection measures must not result in children's exclusion from the digital environment, and that States should address root causes of harm rather than restricting children's access. The Australian ban fails that standard. Jurisdictions considering replication would inherit its defects rather than solve them, and our analysis supplies the doctrinal toolkit for identifying those defects before legislation is passed.

References

- Commonwealth of Australia. (2025). *Age assurance technology trial: Final report*. Department of Infrastructure, Transport, Regional Development, Communications and the Arts.
- Consumer Protection Act 68 of 2008 (South Africa).
- Information Regulator (South Africa). (2021). *Guidance note on processing of personal information of children*.
- South African Schools Act 84 of 1996.
- UN Committee on the Rights of the Child. (2021). *General comment No. 25 (2021) on children's rights in relation to the digital environment* (CRC/C/GC/25). <https://www.ohchr.org/en/documents/general->

comments-and-recommendations/general-comment-no-25-2021-childrens-rights-relation

Wall, J. (2008). Human rights in light of childhood. *International Journal of Children's Rights*, 16(4), 523–543. <https://doi.org/10.1163/157181808X312122>

Wall, J. (2010). *Ethics in light of childhood*. Georgetown University Press.

Young-Bruehl, E. (2012). *Childism: Confronting prejudice against children*. Yale University Press.