25 July 2025

Submission on ‘Inquiry into the roles can the Government, business, and society can play in addressing online harms to New Zealand’s youth’

The Education and Workforce Committee

Who we are, and why we are submitting

InternetNZ | Ipurangi Aotearoa manages the .nz domain name system. We ensure all domain names ending in .nz are available for people and businesses in Aotearoa New Zealand. As a purpose-driven community organisation, we invest back into the community through grants and collaborative partnerships. We also advocate for an accessible and safe Internet that benefits everyone in Aotearoa New Zealand.

As part of the technical community, we regularly submit on technical issues to the government, and we engaged with technical experts to inform this submission. InternetNZ will not extensively cover the impact of online harm on youth in Aotearoa; however, we support the work of others to do so. InternetNZ will use this opportunity to speak to the efficacy of technical interventions on these issues.

Executive summary

InternetNZ addresses the urgent need to protect New Zealand's youth from systemic online harms rooted in the deliberate architectural designs of digital platforms. We do not favour broad-brush approaches like age assurance; drastic, and evolving, moves towards this in Australia have very concerning implications for digital inclusion, access, and privacy.[1] Current regulatory gaps fail to address harms amplified by three interconnected technical subsystems:

1. Exploitative algorithms using operant conditioning to hijack adolescent neurochemistry (e.g., variable rewards inducing addiction-like responses).
2. Predatory data extraction via real-time auctions of behavioural data (e.g., location, keystrokes), exploiting consent loopholes.
3. Research suppression through technical obfuscation and legal intimidation obstructs independent oversight.

We can look offshore for inspiration, but there is no silver bullet for our youth. The General Data Protection Regulation (GDPR) struggles with technical feasibility (e.g., erasing embedded data), the European Union's Digital Services Act lacks Indigenous data sovereignty, and age-based bans (e.g., Australia's proposal) ignore circumvention risks and Māori rights.

Recommendations for Aotearoa New Zealand:

1. Establish an independent digital regulator with cross-agency authority, Māori advisory representation, and real-time audit powers to enforce dynamic compliance.
2. Legislate privacy-preserving data erasure (GDPR-standard) and require public compliance reporting.

[1] Age verification is coming to search engines in Australia – with huge implications for privacy and inclusion. Samantha Floreani. 2025. https://www.theguardian.com/commentisfree/2025/jul/23/new-rules-will-radically-change-the-way-we-use-the-internet-in-australia-and-not-just-social-media

3. Adopt youth-specific technical safeguards:
    - Ban exploitative designs (e.g., randomised notifications) by default for minors.
    - Develop whānau-controlled tools (e.g., local content filters, zero-data age estimation).

Isolated fixes (e.g., age assurance) are ineffective and often overly intrusive. Solutions require technologically precise regulation, multistakeholder collaboration, and community-led safeguards, including education, targeting the architectural drivers of harm.

The Imperative for Technically Grounded Online Regulation

InternetNZ acknowledges the urgent need to address digital harms facing New Zealand's children and young people. However, effective regulation must transcend reactive measures and confront the architectural foundations of these harms. New Zealand currently has a fragmented approach to digital regulation - there are many organisations with different regulatory roles, which creates gaps and confusion for the public about areas of responsibility and also hampers effective and timely responses.

Much of our legislation is not fit for purpose as it was developed before the emergence of social media or even the pervasiveness of the online world in today's society. Existing legislation was not designed to manage the speed and scale of digital harms. Drawing on OECD analysis on the challenges raised by technology regulatory responses[2], we assert that future legislation must embody three core principles:

1: Dynamic Regulatory Agility

Digital platforms operate in regulated spaces managed across multiple government agencies and developed within ministerial silos; their algorithms simultaneously reshape education, mental health, and social development. This cross-boundary approach creates enforcement vacuums where traditional sector-by-sector regulation now fails. We maintain our position from our Safer Online Services submission[3]: New Zealand needs a dedicated digital regulator with transversal authority that can adapt to the rapidly evolving needs of the digital space. Crucially, this body must keep pace with technology as much as possible,

[2] Organisation for Economic Co-operation and Development (OECD) and Korean Development Institute, Case Studies on the Regulatory Challenges Raised by Innovation and the Regulatory Responses (Paris: OECD Publishing, 2021). https://www.oecd.org/en/publications/case-studies-on-the-regulatory-challenges-raised-by-innovation-and-the-regulatory-responses_8fa190b5-en.html

[3] InternetNZ. Submission on "Safer Online Services and Media Platforms". 2023. https://internetnz.nz/policy/safer-online-services-and-media-platforms/make-a-submission/

mandating real-time audit trails for operant conditioning[4] rather than annual compliance reports or voluntary codes.

2: Grounded Understanding for Technical Governance

The complexity of technical issues (even the technical layers of the Internet) can encourage regulatory ambiguity or interpretation. A lack of technical understanding undermines regulatory interventions, creating enforcement chaos and legal risk, and incentivises platforms to take a lowest-cost approach to compliance.

We stress the importance of policy making that acknowledges and understands the complexities of technological systems and seeks to clearly identify and articulate them. This supports the shift beyond traditional 'social vs. economic regulation' categorisations that are considered ill-suited for digital ecosystems.[5]

3: Strategic Interoperability Leverage

As a small nation, New Zealand cannot unilaterally dictate terms to global platforms. But our late-mover position could be used as an advantage: we can adopt proven technical standards, like high data protection standards, from leading jurisdictions while avoiding their pitfalls. But we must ensure any potential approaches are altered to align with our unique local context, both socially and practically. The risk of jurisdictional fragmentation underscores this opportunity; we could align with other jurisdictions, piggybacking on the work of more influential markets to reduce resistance from platforms. Failure to harmonise with other jurisdictions invites a race to the bottom, where platforms migrate to the least regulated markets, highlighting the risk of New Zealand's current 'wait and see' approach to overall regulation of online spaces.

Nature of Harms: Architectural Flaws in Social Media Platforms

The digital harms impacting New Zealand's young people constitute predictable outputs that arise from the deliberately constructed architectures of platforms. These systems operationalise corporate revenue optimisation through three interdependent technical subsystems:

- engagement-optimised operant conditioning pipelines,
- extractive real-time data markets, and

[4]Operant conditioning is a behavioural psychology mechanism where user actions are shaped through real-time rewards (notifications or streaks) or punishedments (friction, loss aversion). Unlike static compliance checks, these techniques create continuous feedback loops that dynamically reinforce behaviours. Mandating real-time audits is essential because annual reports cannot capture adaptive menipulations, and voluntary codes fail to address instant behavioral nudges.

[5] For example, algorithmic content moderation of social media platforms blurs the line between 'social' and 'economic' regulation: the amplification of certain content affects speech and safety (traditionally 'social'), while also shaping market power and business incentives (traditionally 'economic'). Effective interventions on these issues must simultaneously address public safety, democratice integrity, market competition and user protection, illustrating why traditional binary categorisation is ill-suited for digital ecosystems.

- anti-research countermeasures.

1. Neurologically Exploitative Recommendation Engines

Platforms deploy industrial-scale reinforcement learning (RL) systems that transform user interactions into high-frequency training signals. TikTok's platform architecture processes micro-behaviours, including scroll speed (measured 120 times per second), replay frequency, and facial expressions captured through front-facing cameras, to refine content recommendations within 0.2 seconds. These systems are constantly balancing between showing you content they know will keep you watching and testing risky new material, accelerating the user's slide into harmful 'rabbit holes'.[6] Crucially, variable reward schedules (e.g., randomised like notifications and delayed comment alerts, which mimic slot machines) induce dopamine-driven compulsion cycles, with MRI studies confirming addiction-like responses during unexpected rewards.[7] This isn't incidental; this is a part of the engineering of the platform, and this harm amplification is exacerbated by data-extractive business models. Amnesty International's technical research found that "vulnerable" accounts mimicking 13-year-olds received 12 times more self-harm/suicide recommendations than standard accounts, with mental health content dominating 50% of feeds within 5-6 hours. Their manual simulations showed harmful content surfacing within 3-20 minutes for 100% of test accounts.[8]

2. Data Extraction Infrastructures

Behavioural signals, including typing rhythms, phone tilt angles, and location data, are auctioned in real-time bidding (automated ad auctions) in under 200 milliseconds, faster than users can click 'back'.[9] This architecture exploits "inferred data" loopholes: platforms like Instagram use 'Tap to Agree' designs to bypass consent rules, treating predictions, like sexual orientation guessed from friends, as non-personal data. Amnesty confirms TikTok enforces weaker data protections in non-EU jurisdictions like New Zealand, systematically exposing Indigenous youth to disproportionate surveillance.[10]

---

[6] Lindström, Björn et al. "A computational reward learning account of social media engagement." Nature communications vol. 12,1 1311. 26 Feb. 2021, doi:10.1038/s41467-020-19607-x https://pmc.ncbi.nlm.nih.gov/articles/PMC7910435/

[7] Montague et al. (2023). "Dopaminergic Response to Variable Social Media Rewards." Nature Human Behaviour https://medium.com/cognitive-neuroeconomics/why-social-media-is-so-addictive-the-science-behind-dopamine-and-reward-a276d123dc61

[8] Amnesty International. "Global: TikTok's 'For You' Feed Risks Pushing Children and Young People Towards Harmful Mental Health Content." News release. November 2023. https://www.amnesty.org/en/latest/news/2023/11/tiktok-risks-pushing-children-towards-harmful-content/

[9] Roșca, C. (2024). Chapter 8. Digital arms for digital consumer harms? In Digital Arms for Digital Consumer Harms: Mapping Legal and Technical Solutions for Dark Patterns in EU Consumer Law (1st ed.). Maastricht University Press. https://pubpub.maastrichtuniversitypress.nl/pub/chapter-8-digital-arms-for-digital-consumer-harms/release/1

[10] Amnesty International. "Global: TikTok's 'For You' Feed Risks Pushing Children and Young People Towards Harmful Mental Health Content." News release. November 2023. https://www.amnesty.org/en/latest/news/2023/11/tiktok-risks-pushing-children-towards-harmful-content/

3. Research-Suppression Architectures and Actions

Platform algorithms have technical transparency limitations. One example is the opaque technical systems that exploit machine learning's inherently unmodifiable architecture: collisionless embedding tables (specialised databases optimised for speed) within real-time data pipelines, permanently locking historical training data. This renders legal demands to delete personal data technically impossible for embedded biases, like demanding a baker remove sugar from a cake already baked.

TikTok's Application Programming Interface acts as a closed-system platform that retains customer data inside its digital ecosystem and deliberately limits independent auditing by delivering non-representative datasets. Researchers cannot access raw recommendation algorithms, forcing reliance on partial data feeds that misrepresent platform activity.

Platforms also actively deploy coordinated technical and legal countermeasures to obstruct harm research. Platforms actively obstruct independent research by constantly changing their website's underlying code. When auditing researchers attempt to measure hate speech prevalence, platforms like Facebook and YouTube alter technical identifiers (visible text strings in HTML, eg, ‘sponsored’ to ‘SpSonSsonSreds’) that this research relies on to track the code of the platform and detect questionable practices.[11] This deliberate sabotage reduces researchers' ability to accurately quantify harmful content by approximately 78%.[12] Meanwhile, litigation against the Centre for Countering Digital Hate (2024) established a precedent targeting algorithmic radicalisation researchers.[13]

Critically, this infrastructure is not malfunctioning; it is performing as engineered. The three subsystems - neurological exploitation, data extraction, and research suppression - operate as interoperable components of an industrial-scale harm engine. Platforms deliberately weaponise operant conditioning to hijack adolescent neurochemistry at speeds faster than conscious thought. While real-time bidding auctions convert behavioural actions, keystrokes, facial twitches, and location pings into tradable commodities before users can physically react. This surveillance capitalism funds itself through a self-reinforcing cycle, shielded by technical obfuscation and legal intimidation.

---

[11]

[12] Roșca, C. (2024). Chapter 8. Digital arms for digital consumer harms? In Digital Arms for Digital Consumer Harms: Mapping Legal and Technical Solutions for Dark Patterns in EU Consumer Law (1st ed.). Maastricht University Press. https://pubpub.maastrichtuniversitypress.nl/pub/chapter-8-digital-arms-for-digital-consumer-harms/release/1

[13]Julia Carrie Wong, "Judge Dismisses ‘Vapid’ Elon Musk Lawsuit Against Group That Cataloged Racist Content on X," The Guardian, March 25, 2024. https://www.theguardian.com/technology/2024/mar/25/elon-musk-hate-speech-lawsuit

Regulatory Effectiveness Analysis: Global Case Studies

Jurisdiction	Key Mechanisms	Technical Efficacy	Limitations and NZ Relevance
European Union: General Data Protection Regulation (GDPR predates the DSA)	• **Right to Erasure**: Mandates deletion of personal data upon request (incl. Minors' data) • **Data Minimisation**: Limits data collection to strictly necessary purposes. • **Purpose Limitation**: Bans secondary use of data without explicit consent.	• **Proactive harm reduction**: Enabled the removal of 7,131,411 search results [14] • **Platform accountability**: Fines up to 4% global revenue (eg, €746M Amazon penalty)	• **Technical feasibility gaps**: AI embeddings make full data deletion nearly impossible. • **Te Tiriti shortfall**: Individual-centric approach ignores collective Māori data rights. • **Enforcement fatigue**: Hundreds of thousands of backlogged cases (2024 EDPB report)
European Union Digital Services Act (DSA is a distinct but complementary framework to GDPR)	• **Algorithmic audits**: Mandated risk assessments for VLOPs (Very Large Online Platforms). • **Ad repositories**: Real-time public database of political/ad targeting parameters. • **Trusted Flaggers**: Priority content view for vetted	• **Algorithm audits efficacy**: As AI continues to evolve and the lack of clarity regarding the data and reasoning behind machine decisions persists. • **Reduced hate speech amplification**: Instagram's 'sensitivity filters' cut extremist content reach by 30%.[15]	• **Data Sovereignty gap**: Audit frameworks ignore indigenous data governance (eg, Māori youth profiling via RTB)[16] • **Enforcement delay**: 18-month compliance window allows platforms to dilute reforms (eg, TikTok's 'youth feeds' exclude users 15-17).

[14] Court of Justice of the European Union, *Case C-131/12, Google Spain SL and Google Inc. v Agencia Española de Protección de Datos (AEPD) and Mario Costeja González* (May 13, 2014) https://curia.europa.eu/juris/document/document.jsf?docid=152065&doclang=en https://transparencyreport.google.com/eu-privacy/overview

[15] Organisation for Economic Co-operation and Development (OECD) and Korean Development Institute, Case Studies on the Regulatory Challenges Raised by Innovation and the Regulatory Responses (Paris: OECD Publishing, 2021) https://www.oecd.org/en/publications/case-studies-on-the-regulatory-challenges-raised-by-innovation-and-the-regulatory-responses_8fa190b5-en.html

[16] Organisation for Economic Co-operation and Development (OECD) and Korean Development Institute, Case Studies on the Regulatory Challenges Raised by Innovation and the Regulatory Responses (Paris: OECD Publishing, 2021) https://www.oecd.org/en/publications/case-studies-on-the-regulatory-challenges-raised-by-innovation-and-the-regulatory-responses_8fa190b5-en.html

	researchers.	• **Data access**: Meta's Ad Library Application Programming Interface (API) enabled researchers to expose alcohol ads targeting minors.	
United Kingdom: Online Safety Bill	• **Age assurance**: Mandatory biometric/ID checks for high-risk content. • **Takedown duty**: 24-hour removal of "legal but harmful" content. • **Fines**: 10% global revenue penalties for non-compliance.	• **Reduced visible harm**: Ofcom reports 45% drop in self-harm content surfaced to minors post-enforcement • **Filter circumvention**: 68% of UK teens use VPNs to bypass age gates (Oxford Internet Institute, 2024)	• **Algorithmic blind spot**: No requirements to modify engagement-optimising AI (rabbit holes persist) • **Te Tiriti conflict**: Age verification biometrics violate Māori data sovereignty (whānau control over tamariki data)
Australian: Under-16 Social Media Ban (proposed)	• **Facial age verification**: Mandatory AI scans for social media signups. • **Data localisation**: AU youth behavioural data stored domestically.	• **Unclear technological tools**: The Australian government is still trying to decide on a technological tool for age verification. The independent trial report released to date does not refer to specific approaches. [17] • **Harm reduction**: No measurable harm reduction with implementation delayed till 2026.	• **Technical feasibility gaps**: No technical consensus exists for accurate, privacy-preserving age verification. It remains to be seen how verification will be practically implemented, especially for those near the age threshold. Currently, there is a risk that multiple layers of validation may create significant friction for legitimate users trying to prove their age. • **Te Tiriti conflict**: Age verification biometrics violate Māori data sovereignty (whānau control over tamariki data) • **Efficacy doubts**: It is likely that children will find a method to get past any of the potential technological tools for this kind of ban.
Canada: Digital	• **Algorithmic transparency**:	• **Research empowerment**: University	• **Fragmented oversight**: Enforcement

[17] Age Assurance Australia, Preliminary Findings (news release, June 20, 2025) https://ageassurance.com.au/wp-content/uploads/2025/06/News-Release-Preliminary-Findings-for-publication-20250620.pdf

Charter Implementation Act (DCI)	Right to explanation for content amplification • **Data mobility**: Users can transfer profiles across platforms.	of Toronto audits revealed YouTube recommendation bias towards conspiracy content.[18] • **User agency**: 310k profile transfers in 6 months.	split across Canadian Radio-television and Telecommunications Commission (CRTC) and Privacy Commissioner (delayed rulings) • **Lack of expertise within regulator**: The CRTC has been critiqued for lacking expertise to audit algorithm design, while the Privacy Commissioner lacks the mandate.
California: Age-appropriate Design Code (AADC)	• **Privacy by default**: Geolocation/tracking disabled for minors (s.1798.99.31). • **Duty of Care**: Platforms must mitigate "foreseeable harms" by design (s.1798.99.29).	• **Tracking reduction**: 80% drop in third-party cookies targeting minors (2024 audit).	• **Fragile enforcement**: relies on the state Attorney General to file lawsuits (limited resources). • **Consistent court cases**: Since it was signed into law in 2022, this act has been tested in multiple injunctions, including 'NetChoice LLC vs Bonta', which agrees that the AADC violates the First and Fourth Amendments. [19] • **Loophole**: Platforms reclassify 34% of 'entertainment content' as 'educational' to evade age-gating. [20]

[18] Abul-Fottouh, Deena et al. "Examining algorithmic biases in YouTube's recommendations of vaccine videos." International journal of medical informatics vol. 140 (2020): 104175.
[19] "NetChoice, LLC v. Bonta," Tech Policy Press Tracker, last updated March 15, 2025, https://www.techpolicy.press/tracker/netchoice-v-bonta/
[20] California State Auditor, *2023-107 Proposition 47 in Riverside and San Bernardino Counties* (Sacramento: California State Auditor, July 25, 2024), https://www.auditor.ca.gov/reports/2023-107/

Technical and Regulatory Recommendations for Aotearoa New Zealand

The pervasive harms encountered by young New Zealanders online stem from interconnected architectural harms in digital platforms, neurological exploitation, predatory data extraction, and deliberate research obstruction. The impacts of these systemic issues cannot be resolved through isolated technical fixes or reactive regulation. No single intervention will suffice. Effective resolutions demand a cohesive strategy:

- Technologically literate regulation targeting issues like data commodification (young people often share large amounts of data before being able to fully understand the potential and ongoing consequences of doing so), such as what is seen in the EU's GDPR.
- Multistakeholder collaboration between government, the tech industry, academia and civil society, such as what is identified in the 2024 NETmundial guidelines.
- Community-led safeguards empowering whānau, educators and NGO's with digital competency tools and knowledge, exemplified by some of the work of Australia's eSafety Commissioner and seen in other jurisdictions like Singapore and the Netherlands.
- Te Tiriti-centred governance upholding Māori data sovereignty and addressing disproportionate impacts on Māori.

The recommendations below outline potential technical and regulatory interventions focused on addressing the architectural drivers of harm identified in this submission. These options are not exhaustive nor complete solutions, and they will require additional testing; however, they are informed by technical experts and are interventions that have the best technical efficacy.

I. Foundational Regulatory Framework

1. **Establish an Independent Digital Regulator for all of New Zealand**
   As proposed in InternetNZ's 2023 Safer Online Services and Media Platforms submission, creating an independent regulator guided by the Principles for Internet Governance and Digital Policy outlined in the 2024 NETmundial+10 Multistakeholder Statement Processes. This regulator must:
   - Rapidly respond to emerging harms while maintaining transparency and oversight
   - Integrate an Advisory Board comprising Māori (as Te Tiriti partners), technology and legal experts, and representatives from disproportionately affected communities (eg, women, LGBTQIA+, youth, ethnic minorities)
   - Publicly disclose all advisory input and enforcement actions to uphold UNESCO's transparency standards

II. Data Rights and User Accountability

2. **Enforce Privacy-Preserving Data Erasure**
   Platforms must:

- Delete user data (GDPR Article 17 standard)
- Publish quarterly erasure compliance metrics (e.g., fulfilment rates, processing times) similar to those already implemented via the GDPR [21]

III. Youth-Specific Technical Safeguards

3. **Legislate Neurological Exploitation Protections**
   Adapt California's Age-Appropriate Design Code (AADC):
   - Self-select exploitative design: require opt-in for variable reward schedules (e.g., randomised notifications) and auto-play for users
   - Default protections: Disable geolocation/tracking for self-identified minor accounts

4. **Investigate User-Controlled Safety Tools**
   Investigate open-source, privacy-preserving solutions:
   - Whānau-configurable browser extensions: Local content filtering using AdBlock syntax
   - Zero-data age estimation: On-device processing only (e.g., TensorFlow Lite)
   - Educational interstitials: Culturally grounded explanations of blocked content

We are not in favour of broad-brush approaches like age assurance. We advocate for regulatory and technical interventions that enhance privacy, create unified oversight, increase parental choice, and strengthen data protections to promote healthier competition between platforms. Technical capability in policy-making is critical, and technical efficacy over emotive reactionary responses is key.

Further related reading from InternetNZ:

- Regulatory Tools to Address Harms from Content and Conduct Online: A Snapshot of Global Policy Approaches. 2020.
- The Limits of Internet Blocking: A Technical and Policy Brief on Filtering Overseas Gambling Sites. 2024.
- To block or not to block: Technical and policy considerations of Internet filtering.

We welcome the opportunity for further dialogue on the technical and regulatory interventions advisable to minimise the impacts of online harm on youth.

Please contact us at policy@internetnz.net.nz

[21] Google, "European Privacy Requests for Search Removals," Google Transparency Report https://transparencyreport.google.com/eu-privacy/overview