



Regulatory Impact Statement: Reducing the risk of online harm to children: Phase One Policy Decisions

Decision sought	This Regulatory Impact Statement (RIS) supports final Cabinet decisions on the policy design for Phase One of a two-step regulatory approach to reduce the risk of online harm to children.
Agency responsible	Internal Affairs
Proposing Ministers	Hon Erica Stanford, Minister of Education
Date finalised	11 March 2026

Description of the Minister's regulatory proposal:

Cabinet agreed in December 2025 to develop a two-step plan to improve children's online safety, including the broad direction for both phases [SOU-25-MIN-0177 refers]. For Phase One, Cabinet agreed at a high level to develop options for a minimum age of 16 for social media and other accounts, and to require platforms to assess and report on risks of harm to all children under 18. The Minister of Education now seeks Cabinet agreement to the final policy design decisions needed to enable drafting of the Phase One legislation, including:

- the scope of the platforms that will be regulated;
- the age-assurance and privacy requirements platforms must meet;
- the content and process for mandatory child-safety risk assessments and reporting; and
- the functions, powers, and enforcement tools required for the Phase One regulator to oversee compliance.

Phase Two will consider options on a broader online safety framework, 9(2)(f)(iv)

Both phases will be complemented by non-regulatory educational initiatives.

Summary: Problem definition and options

What is the policy problem?

Children and young people in Aotearoa are exposed to a wide range of online harms, including harmful content (violent, self-harm, sexualised, body-image-related, dangerous challenges) and harmful behaviours (cyberbullying, grooming, unwanted contact, interactions with Artificial Intelligence). Some platform design features, such as a design centred around endless scrolling and feeds that automatically recommend targeted content, can compound these risks and contribute to problematic use.

Social media and online platforms (platforms) provide benefits, including social connection, access to information and wellbeing support. These can be particularly important for Māori, Pacific, LGBTQIA+, disabled and rural youth. However, some these groups also experience disproportionately higher levels of online harm, increasing the need for targeted, effective protections.

New Zealand's current regulatory settings are outdated and fragmented, focused on content and individual conduct, and are reliant on voluntary platform action and self-reporting by users. Platforms have limited regulatory incentives to prioritise child safety, and the existing framework does not require proactive risk mitigation or safe design.

Government intervention is required to address these gaps. The Government has committed to considering a minimum age for social media access, and the Prime Minister has committed to introducing a Bill before the 2026 election. Global momentum for stronger online safety regulation (e.g., EU, UK, US, Australia) reinforces the need for modernised protections.

What is the policy objective?

The primary objective of this work is to promote the protection of children from the risk of exposure to online harm, particularly under 16s. Phase One would advance this objective by:

- ensuring under-16s cannot create or hold accounts on social media and other platforms
- requiring platforms to assess, mitigate and report risks to all under-18s
- establishing regulatory capability ahead of wider reforms in Phase Two.

Success indicators include:

- reduced rates of cyberbullying, grooming, and exposure to harmful content
- improved wellbeing indicators
- compliance with age assurance requirements by platforms
- reductions in complaints to regulators, platforms and organisations like Netsafe.

What policy options have been considered, including any alternatives to regulation?

Primary intervention options considered:

- Option 1 (Preferred): Counterfactual – Minimum age of 16 + and mandatory child safety reporting (Phase One), with Phase Two to develop a broader online safety framework.
- Option 2: Allow platforms to create 'teen accounts' for underage users.
- Option 3: Rely on current voluntary reporting by platforms.
- Option 4: Require platforms to have a duty of care to protect children.

Design question one: Scope of platforms

- Option 1: Replicate Australian purpose-based social media scope.
- Option 2 (Preferred): Australia + social AI companions.
- Option 3: Option 2 + online multiplayer videogames.

Design question two: How platforms comply with minimum-age requirements.

- Option 1 (Preferred): High-level, outcomes-based "reasonable steps".
- Option 2: Prescriptive requirements.

Design question three: What platforms should report on.

- Option 1 (Preferred): Comprehensive statutory list of risks.
- Option 2: Short list aligned with voluntary Code of Practice categories.

Design question four: How the regulator should penalise serious non-compliance.

- Option 1 (Preferred): Financial penalties and business disruption tools, including service disruption orders (which can require ancillary services to stop servicing a non-compliant platform) and access restriction orders (which can block access to a platform in New Zealand).
- Option 2: Financial penalties only.

Non-regulatory approaches such as education, awareness-raising and parental controls on apps are insufficient on their own. These tools are important for building public understanding and should complement, rather than replace, regulatory measures.

What consultation has been undertaken?

Early engagement occurred with domestic and international regulators and child safety organisations, including the Australian Government and eSafety Commissioner, Ofcom (UK), Coimisiún na Meán (Ireland), Classification Office, Netsafe, Office of the Privacy Commissioner, Human Rights Commission, Mana Mokopuna and age assurance/digital identity technology providers. Policy development has also been informed by the Education and Workforce Committee's interim and final reports from its inquiry into the harm young New Zealanders encounter online.

Targeted consultation was carried out from 21 January to 5 February. Engagement covered 75 stakeholders, including major platforms, industry bodies, age assurance and digital identity experts, youth (aged 12–20), independent Crown Entities, Māori and population specific organisations (including disabled, rainbow, Pacific and rural groups), medical and child mental health experts, international regulators, child rights advocates, and wider civil society. Appendix A includes a full list of the stakeholders included in consultation. Feedback was sought on key design features of the proposed Phase One approach, including age assurance requirements, platform obligations, scope, and regulatory design. If Cabinet agrees to progress this legislation, broader public consultation will occur through the select committee process.

Key themes from our targeted consultation include:

- Youth views were mixed. Many recognise online-harm risks and support safer design and an age restriction. Others oppose blanket age limits and emphasise keeping access to connection and support, especially for marginalised groups.
- Parents, educators and child-rights advocates support stronger regulation. Some questioned the sequencing of phases, preferring a duty of care and stronger rules for platforms before a minimum age is considered. They also highlighted the importance of education, digital literacy and family focused support.
- Industry acknowledges the need for improved safety but emphasises privacy, proportionality, technology neutral rules, and realistic implementation timeframes.
- Many groups highlighted social AI companions as a safety concern and supported their inclusion into the scope of Phase One.
- There were significant concerns about age-assurance accuracy, privacy and data risks, exclusion of young people without ID, poor biometric performance for Māori and Pacific youth, circumvention, and the risk of pushing children into less safe spaces.
- There was broad support overall for a well-resourced regulator with robust enforcement powers.

Is the preferred option in the Cabinet paper the same as preferred option in the RIS?

Yes

Summary: Minister's preferred option in the Cabinet paper

Costs (Core information)

Key costs relative to the counterfactual:

- **Children under 16:** loss of logged in access; potential displacement to unregulated services; unquantifiable social connection impacts.
- **Platforms:** setup and ongoing costs for age assurance, NZ specific risk reporting, and governance systems; higher burden for smaller services.
- **Government:** establishment and operation of the Phase One regulator [BUDGET SENSITIVE] (~\$6–8.5m p.a. across ramp up period, including policy FTE funding).
- **Parents/caregivers:** minor time and communication costs.
- **Wider public over the age of 16:** minimal effort or inconvenience for creating a new account; medium privacy/trust considerations.

Benefits (Core information)

Key benefits relative to the counterfactual:

- **Children under 16:** Reduced exposure to harmful content, harmful contact, design features which can contribute to higher risk of harm, and problematic use patterns.
- **All children under 18:** Safer digital environments because of platform level obligations.
- **High-risk groups:** Reduced disproportionate exposure to harm for Māori, Pacific, LGBTQIA+, and disabled youth.
- **Regulators and government:** Improved visibility of platform risks; evidence for Phase Two; alignment with international practice.
- **New Zealand's online ecosystem:** Stronger public confidence in platform responsibility; alignment with international online safety frameworks.

Balance of benefits and costs (Core information)

Does the RIS indicate that the benefits of the Minister's preferred option are likely to outweigh the costs?

Although full monetised analysis is not available, qualitative and early quantitative evidence strongly indicates benefits outweigh costs, particularly for children and families experiencing harm. Benefits are expected to increase over time as platforms embed safer design, and as risk-mitigation reporting requirements become normalised. A range of complementary assessments (child-impact, privacy-impact, Te Tiriti, human rights) have informed the policy design. Remaining uncertainties relate to implementation costs for platforms, data gaps on harm prevalence, and platform behavioural responses. Continued consultation with regulators and experts will further refine assumptions and proposal details but do not prevent Cabinet making an informed decision.

Implementation

How will the proposal be implemented, who will implement it, and what are the risks?

Legislation is intended to be introduced before the end of the parliamentary term. The Phase One regulator will administer the framework, monitor compliance, provide guidance, and take enforcement action. It will also have a role in educating children, their caregivers and the wider public on the Phase One changes.

Implementation risks and mitigations include:

1. **Circumvention by children:** Some level of circumvention is unavoidable, but this risk is mitigated through layered, privacy preserving age assurance expectations and requirements for platforms to monitor and respond to circumvention methods. No penalties on young people or parents avoids discouraging help seeking.
2. **Shift to alternative or fringe services:** Managed through education, awareness, and careful scoping of regulated platforms.
3. **Privacy concerns about age assurance:** Addressed through strict limits on data collection, transparency, data minimisation, and compliance with the Privacy Act 2020.
4. **Platform withdrawal or service disruption:** Mitigated through strong alignment with Australia and proportionate regulatory requirements.
5. **Equity risks:** Addressed through child impact and Te Tiriti assessments and targeted engagement with high-risk communities.

Limitations and Constraints on Analysis

Policy development for Phase One was constrained by Cabinet's direction to develop a minimum age and reporting regime, alongside compressed timeframes, the need to minimise Crown expenditure, limited New Zealand specific data, limited international experience with minimum age regimes, and the need for practical alignment with Australia. Despite these constraints, the analysis is robust enough to support Cabinet decisions on the final design of Phase One.

I have read the Regulatory Impact Statement and I am satisfied that, given the available evidence, it represents a reasonable view of the likely costs, benefits and impact of the preferred option.

Responsible Manager(s) signature: 9(2)(a) _____
9(2)(a)

11 March 2026

Quality Assurance Statement <i>[Note this isn't included in the four-page limit]</i>	
Reviewing Agency: DIA	QA rating: Meets
Panel Comment: The panel considers that the information and analysis summarised in the RIS meets the quality assurance criteria. The RIS contains the necessary information. The financial, scope and policy constraints are clearly identified, and the RIS signals issues what will be considered as part of Phase Two of the reforms. The RIS has a strong evidence base, drawing on the information and research that is available for an issue that is evolving quickly (including relevant international comparisons). The analysis is convincing with clear articulation of the weighting and trade-offs for the criteria. Consultation has been undertaken across a broad range of interests and reflected in the analysis. The paper strikes a good balance between presenting enough supporting information and being easy to read.	

Section 1: Diagnosing the policy problem

What is the context behind the policy problem and how is the status quo expected to develop?

1. In December 2025, Cabinet agreed to develop a two-step plan to improve children's online safety. As the first phase, Cabinet agreed to develop options for a minimum age requirement of 16 for social media and other accounts and options for those platforms to report on the risks of harm to all children under 18 [SOU-25-MIN-0177 refers]. An [Interim Regulatory Impact Statement](#) was prepared to inform these decisions, and a public link to this document is intended to be available pending Cabinet decisions. Cabinet has invited the Minister of Education to report back on final policy design in March 2026. This Regulatory Impact Statement provides the detailed analysis needed to support Cabinet's decisions on the final Phase One design and addresses the gaps identified in the interim regulatory impact statement.
2. New Zealand's current regulatory framework does not adequately safeguard children (under 18s) online, and critically, it overlooks the heightened developmental vulnerability of children aged under 16. While some protections exist, there is no overarching framework that regulates social media platforms or other platforms where 'social' interaction (either with other users or Artificial Intelligence) occurs to prevent risk of exposure to harm.
3. New Zealand's current system for reducing the risk of harm from online content and activity includes:
 - The Films, Videos, and Publications Classification Act 1993 (FVPCA) is a key lever in the regulatory system. The Act has two mechanisms intended to ensure consumer safety; objectionable material which is illegal to possess or trade, and content that is legal to

create, possess and distribute, subject to access restrictions such as R16 and R18 labels. It does not provide a mandate to regulate systems of online platforms in promoting content, conduct or features, some of which is harmful, explicit or illegal.

- The Harmful Digital Communications Act 2015 (HDCA), which aims to deter, prevent and mitigate harm caused by digital communication between individuals. It provides a complaints and redress mechanism for victims of online bullying, harassment, and other interpersonal harms. It does not require platforms to prevent harm or ensure safe design, and its high evidence threshold can limit its ability to address systemic risks.
 - The Unsolicited Electronic Messages Act 2007 (UEMA) focuses solely on preventing spam by restricting unsolicited commercial electronic messages and requiring consent and unsubscribe mechanisms.
 - Voluntary self-regulation by operators, including through schemes managed by the New Zealand Media Council, Netsafe and NZ Tech. For example, the Aotearoa New Zealand Code of Practice for Online Safety and Harms 2022 addresses areas such as bullying, child safety, disinformation, harassment, hate speech and misinformation with a mechanism for lodging complaints against breaches.
4. In March 2026, the Education and Workforce Committee (EWC) delivered its final report of an inquiry into the online harms experienced by young New Zealanders and the roles that government, industry, and communities should play in addressing them. The Committee's report considers that harm to young New Zealanders from online platforms is severe and requires urgent responses from Government, business and society alike. Its overarching conclusion is that the Government should consider ways to improve institutional frameworks, education, and legislation underpinning the regulation of online safety, and take further steps to support parents and caregivers in protecting their young people.¹
 5. The report made 12 recommendations for the Government to improve online safety including restricting social media access for under-16-year-olds, establishing an independent national regulator for online safety, regulating algorithmic recommendation systems and strengthen online platform's civil and regulatory liability for harm resulting from the content they host and their platform's design.
 6. In 2023, the Department of Internal Affairs (DIA), with support from the Ministry of Culture and Heritage (MCH), consulted on proposals to review the media and online content system.² This was in response to concerns about rising online harms, outdated legislation, and the inconsistent approach to dealing with harmful content. This work has now concluded.
 7. Three Member's Bills were put forward in 2025 that seek to address gaps in New Zealand's online safety settings:
 - Catherine Wedd MP's Members' Bill to protect under 16s from social media harm is proposing a social media age restriction for under 16s;
 - Laura McClure MP's Deepfake Digital Harm and Exploitation Bill, which strengthens protections against the non-consensual creation and distribution of intimate synthetic imagery; and

¹ Education and Workforce Committee, [Inquiry into the harm young New Zealanders encounter online, and the roles that Government, business, and society should play in addressing those harms](#), March 2026.

² Department of Internal Affairs, [Safer Online Services and Media Platforms](#), June 2023.

- Reuben Davidson MP's Online Safety (Duties for Providers of Internet-Based Services) Bill seeks to hold online platforms accountable for harmful content.
8. Both Catherine Wedd MP and Laura McClure MP's Members' Bills were introduced following their draw from the ballot on 23 October 2025.

How is the status quo expected to develop if no action is taken?

9. If no action is taken, children's exposure to harmful content, contact and design driven- risks is expected to continue. Research shows that 24 percent of children aged 5–14 use social media, with usage rising sharply from age 9.³ The Classification Office reports widespread concern about children's access to illegal or inappropriate content across social media and other online platforms.⁴ New Zealand Police also report an average of ten sextortion cases each week, with 18 percent of victims under 17.⁵ Together, these indicators show that without intervention, children will remain exposed to environments where harmful content and contact can occur.
10. Emerging harms associated with AI-driven features are also expected to continue without regulatory intervention. AI companions are increasingly embedded in major platforms and can create dependency and inappropriate reinforcement. Without intervention, under-16s will continue to be exposed to these harms and reliance on voluntary platform action will remain. The risk of children moving into unregulated environments would also continue.

International context

11. Online safety is a global issue, and a range of countries have looked to progress several different regulatory approaches. Typically, the strongest protections are designed for children, but some countries have developed protections which apply to all users. At this stage, minimum age requirements for online platforms have not been introduced in isolation and are typically part of a broader online safety system.
12. Aligning with international approaches lowers compliance burdens for platforms that operate globally, supports more consistent enforcement across jurisdictions, and allows New Zealand to build on emerging international practice so children receive protections that are broadly consistent with comparable countries.

The Australian approach

13. The Australian approach to a minimum age is designed to reduce harms associated with logged-in use of social media, including exposure to harmful or misleading content, risks to privacy and wellbeing, and the impact of addictive design features such as the ability to endlessly scroll algorithmic content feeds.
14. The Online Safety Amendment (Social Media Minimum Age) Act (SMMA), passed in December 2024, set a minimum age of 16 for social media accounts. The SMMA Act places responsibility on platforms to enforce age restrictions and implement robust age-assurance measures, with penalties of up to NZD\$58.9 million for non-compliance. Australia's policy is supported by an existing broader regulatory framework that includes the Online Safety Act 2021 and the eSafety Commissioner, an independent regulator with powers to enforce platform accountability and set online safety expectations.

³ New Zealand On Air, [Where Are The Audiences - Children's Media Use Report](#), 2025.

⁴ Classification Office, [What We're Watching: New Zealanders views about what we see on screen and online](#), 2022.

⁵ Police Association, [NZPA: Sextortion: It's not just blackmail](#), 2024.

15. The SMMA took effect on 10 December 2025, and as of mid-January 2026 platforms have reported removing approximately 4.7 million accounts identified as belonging to under-16 users.⁶ Two High Court challenges to the minimum-age requirement (including one brought by Reddit and another supported by the Digital Freedom Project) are underway; no decision dates have been set, and judgments are not expected before late 2026.

16. **Appendix B** summarises online safety initiatives in other jurisdictions.

What is the policy problem or opportunity?

Young people can be exposed to harmful content, behaviour, and design features which facilitate the risk of harm on social media and other platforms

17. Social media has become an embedded part of how many children communicate and participate in social life. It has evolved over the years from enabling connections with people we know (e.g. Facebook friends), to connections with unknown people (Twitter/X followers), to an algorithm-led content provision platform with varying levels of direct control.
18. There is no legal definition of “social media” in New Zealand. Colloquially, the term is used to describe online services where users create accounts, share content and interact with others such as TikTok, Instagram, and Snapchat. The boundary is not clear-cut, as many platforms combine social, messaging, gaming, live-streaming and AI-driven features. An analysis of which platforms would likely be in scope of Phase One requirements is set out on pages 22–26.
19. There are positive associations between wellbeing and social media use. Social media has become part of building relationships, forging connections, and staying in touch. As noted in Kōi Tū's evidence brief on the effects of screen time on children's development, “Sharing comments, messages, and “likes” is associated with boosted self-esteem, better-perceived closeness, reduced stress and loneliness, and a more positive outlook. Teenagers can also join online communities which foster belonging and provide support, particularly for minority groups such as LGBTQ+ youth.”⁷
20. Despite the positive associations outlined above, social media is the type of platform where children most recently report experiencing a range of online harms.⁸ A 2025 New Zealand school census asked students under 16 which platforms they had used four or more times the previous day, YouTube led with 51.3 percent of students, followed by TikTok (46.9 percent), Snapchat (43.9 percent) and Instagram (35.3 percent).⁹
21. The World Health Organization's report on the digital determinants of youth mental health and wellbeing found there is strong consensus that children's and adolescents' stages of biological and cognitive development make them particularly vulnerable to online harms.¹⁰ Research referenced by the Australian Government suggests that by 16 years of

⁶ eSafety, [Platforms restrict access to 4.7 million under-16 accounts across Australia](#), January 2026.

⁷ Studies referenced on referenced in the University of Auckland and Kōi Tū: The Centre for Informed Studies, [Screen time: The effects on children's emotional, social and cognitive development](#), 2021, p5.

⁸ eSafety Commissioner, [Digital use and risk: Online platform engagement among children aged 10 to 15](#), 2025.

⁹ Analysis of data provided by [CensusAtSchool](#) on Oct 9, 2025. Includes responses from students aged under 16 from the 2021–22, 2023–24 and 2025 CensusAtSchool.

¹⁰ World Health Organisation. [Addressing the digital determinants of youth mental health and well-being](#), 2025.

age, young people are beyond the most vulnerable stage of adolescence, so restricting access to social media is less likely to be impactful if it does not capture this age group.¹¹

22. Online platforms also use features that increase time spent on them. In 2023, the US Surgeon General's Advisory on Social Media and Youth Mental Health identified several design features that can heighten risk, including push notifications, autoplay, infinite scroll, 'likes' and 'reactions', and algorithmic content delivery based on a user's preferences and interests.¹² Features and functionalities which can increase children's risk of exposure to harmful content have also been recognised by the risk assessments provided to the UK's online safety regulator. Some examples include ability to post images, commenting, searching for user-generated content, and user profiles.¹³
23. A nationally representative study of New Zealand teens aged 14-17 years old published in 2018 found that 70 percent of teens experienced unwanted digital communication (including being contacted by a stranger, and accidentally seeing inappropriate content online), most commonly through social media or through online gaming. 80 percent of those had an emotional response to the unwanted communication, and 19 percent said it had a negative impact on their daily lives.¹⁴
24. More research is also needed to fully understand the causative negative impact of social media on children's health and wellbeing. However, there is a substantial level of research with clear indicators of children being exposed to harmful content and conduct on social media platforms and other platforms where social connection (either with a person or a person-like feature such as an AI companion) happens.
25. A 2022 report by the Classification Office found high levels of concern among New Zealanders about children seeing harmful content, particularly in online spaces. New Zealanders believed online content can negatively influence children, regarded harmful content as a real and widespread problem for tamariki and rangatahi, and supported stronger regulation. Only a minority felt the current system works well to keep young people safe, and just 33 percent had confidence that online platforms provide what people need to stay safe.¹⁵ **Appendix C** contains additional research that further outlines and evidences the policy problem.

Rangatahi Māori and Pacific are more likely to experience harm

26. While New Zealand-specific research is limited, available studies suggest that some groups may be disproportionately affected. Pacific youth are more likely to experience cyberbullying/online social exclusion, receive requests for sexually explicit material, have their personal information posted without permission, and accidentally encounter disturbing material than their peers from other ethnic groups, as well as being more likely to download viruses/malware.¹⁶
27. Rangatahi Māori report spending significantly more time on social media than NZ European youth, and are more likely to be exposed to harmful content online. Māori and Pacific teens are also more likely to report receiving unwanted digital communication that negatively impacted their daily activities compared with other ethnic groups.¹⁵

¹¹ Australian Government, [Social Media Minimum Age – Fact sheet](#), 2025.

¹² The U.S. Surgeon General's Advisory, Social Media and Youth Mental Health, 2023.

¹³ Ofcom, [Year 1 Online Safety Risk Assessments – Report on Standards and Improvements](#), January 2026, p11.

¹⁴ Netsafe, [New Zealand Teens and Digital Harm: Statistical insights into experiences, impact, and response](#), 2018.

¹⁵ Classification Office, [What We're Watching: New Zealanders views about what we see on screen and online](#), 2022.

¹⁶ Journal of Pacific Research, [The impacts of social media on the wellbeing of Pacific youth in New Zealand](#), 2024.

Young people can be exposed to harmful or inappropriate interactions, dependency risks, and unsafe content through AI companions

28. An overview of the current evidence base on the psychological effects of AI chatbots on children provided by the American Psychological Association recommends that AI systems designed for adults are fundamentally inappropriate for youth and require specific, developmentally informed safeguards.¹⁷
29. AI companions are available as standalone platforms and are becoming embedded across major social-media platforms. These systems increasingly sit alongside children's everyday online interactions, offering advice, emotional feedback, and companionship in ways that can blur the line between tool and relationship. Research in 2025 found that conversational AI systems affirm inappropriate or harmful content in 42 percent of cases, reflecting high levels of social sycophancy that can reinforce unsafe thinking or behaviour. A UK survey in 2025 found that 35 percent of children aged 9–17 report that talking to a chatbot feels like talking to a friend, and six in ten parents were worried their children believe these systems are real people.¹⁸ These patterns indicate a growing risk of emotional dependency and displacement of real social support.
30. International regulators in the UK, EU and Australia are also growing in regulatory action against AI chatbots more broadly, as well as AI companions. This is addressing wider issues than the risk of companionship, such as generation of non-consensual imagery or child sexual exploitation material. As part of Australia's wider Online Safety Act powers (separate to minimum age requirements), in October 2025 eSafety required AI companion providers to explain how they are protecting children from exposure to a range of harms, including suicidal ideation and self-harm.¹⁹ eSafety has also noted the risks of AI companions, including sharing harmful content, distorting reality and giving advice that can be dangerous. They are often designed to encourage lengthy interactions and to stimulate ongoing connection.²⁰

New Zealand's content regulatory system is outdated and has gaps

31. New Zealand's content regulatory system currently deals with the most severe and illegal content like child sexual exploitation and promotion of terrorism or violence, regardless of whether it appears online or in traditional forms of media such as printed publications. It also addresses harm between individuals through the HDCA. The system has increasingly become outdated. New Zealand's content regulatory system is over 30 years old and was developed for a media environment dominated by physical media and professional publishers across print, television and film. There are gaps in the current legislative framework which are not sufficiently addressed. For example:
 - The HDCA does not place obligations on online platforms to prevent harm or ensure safe design, and its high evidence threshold limits its ability to address systemic risks;
 - The FVPCA is designed for a pre-digital era, and it does not cover the role of online platforms in curating and promoting harmful content; and

¹⁷ American Psychological Association, [Examining the Harm of AI Chatbots](#), September 2025.

¹⁸ Internet Matters, [Me, myself & AI: Understanding and safeguarding children's use of AI chatbots](#), 2025.

¹⁹ eSafety, [eSafety requires AI companion chatbot providers to explain how they are keeping Aussie kids safe](#), 2025.

²⁰ eSafety, [AI companions: information sheet](#), February 2025.

- The Broadcasting Act 1989 was designed to apply primarily to linear television and radio broadcasting. It is unclear whether the legislation and broadcasting standards could apply to the online environment and this is in the process of being tested.²¹

What objectives are sought in relation to the policy problem?

32. The primary objective of this work is to promote protection of children from the risk of exposure to online harm, particularly to under 16s.

What consultation has been undertaken?

33. Ahead of Cabinet's December 2025 decision to adopt a phased approach to reducing online harm to children [SOU-25-MIN-0177], officials undertook targeted early consultation with international regulators, domestic oversight bodies, and government agencies. Engagement confirmed broad support for stronger protections but highlighted the need for a clearly scoped framework, robust privacy safeguards, and further targeted consultation, including with children and communities likely to be disproportionately affected.
34. Stakeholders raised consistent concerns about the privacy and practicality of age assurance, risks of unintended impacts on low risk services or gaps that leave high risk environments unregulated, and the importance of aligning with existing regulatory settings, including Cabinet's June 2025 decision to take a proportionate and risk-based approach to regulating Artificial Intelligence. Several agencies recommended beginning with a comprehensive framework rather than a minimum age restriction and noted that online harms affect wider user groups, not only children. Agencies also highlighted the importance of considering findings from the Education and Workforce Committee inquiry and cautioned about potential extraterritorial impacts for platforms based overseas.
35. Between 21 January and 5 February 2026, officials carried out targeted consultation to test key design options for Phase One. Engagement covered 75 stakeholders, including major platforms, industry bodies, age assurance and digital identity experts, youth (aged 12–20), Māori and population specific organisations, medical and mental health experts, international regulators, child rights advocates, and wider civil society. **Appendix A** includes a list of the groups that were engaged during targeted consultation. Discussions centred on the scope of platforms subject to minimum age requirements, the obligations they should hold, enforcement options, and the potential for unintended impacts.
36. Stakeholders emphasised significant risks associated with social AI companions and supported their inclusion in Phase One where they can be clearly distinguished from general purpose- chatbots. Participants also called for more prescriptive, New Zealand specific- -child safety reporting requirements, including transparency of methodologies and real world- examples of harm. In response to feedback from the Office of the Privacy Commissioner, officials have proposed more detailed, purpose based data handling- requirements for compliance with age assurance requirements. Feedback also supported the introduction of executive liability for false or misleading information.
37. Youth feedback played an important role in refining the education approach, shifting the focus toward supporting those approaching 16 to make informed choices about how they engage online. Young people also highlighted the need for a layered education strategy,

²¹ The Broadcasting Standards Authority sought independent legal advice in [December 2025](#) on whether the Broadcasting Act could apply to the online environment and brought a complaint against the online-only commentator website 'The Platform'. This complaint is ongoing.

codesigned content, peer and mentor based learning, clearer guidance for whānau, and tools that help young people understand and manage privacy. Across stakeholders, there was strong support for a regulator with both education and enforcement responsibilities, equipped with tools capable of influencing platform behaviour and driving compliance.

38. Across platforms, there was broad support for strengthening online safety but consistent caution about adopting blunt minimum-age bans or prescriptive platform lists. Most platforms argued that harms differ by features rather than platform type, favouring risk-based, safety-by-design approaches over blanket restrictions. They emphasised significant privacy, equity and accuracy concerns with age-assurance technologies, particularly ID-based verification. Platforms also stressed the need for clear problem definition, international alignment, and proportionate, technically feasible obligations that recognise differences between services. Several platforms highlighted risks of unintended consequences, including pushing young people into less safe environments, reducing access to beneficial safety features, or disadvantaging vulnerable groups. They consistently supported a collaborative, technically informed regulator, adequate lead-in time, and a strong public education component.
39. Targeted consultation also provided clearer insight into how a minimum age requirement may affect vulnerable groups, particularly Māori, Pacific, disabled, rural young people, and children and young people in care. These stakeholders noted that under16s may face unintended impacts such as losing access to supportive online spaces, reduced use of safety tools designed for teens, and being diverted into less safe, or unmoderated services. For those aged 16 and over, the primary risks identified relate to age assurance processes, especially where verification relies on formal identification or biometrics, which may disproportionately disadvantage groups less likely to hold formal ID or who face higher privacy concerns. Across all groups, stakeholders emphasised the need for culturally grounded, accessible education, privacy-preserving age assurance options, and strong regulatory safeguards to avoid exacerbating inequities or disconnecting vulnerable young people from essential support networks.

Section 2: Assessing options to address the policy problem

What criteria will be used to compare options to the status quo?

40. The following criteria are applied to evaluate the relative strengths and limitations of each option against the status quo:

- **Certainty** The public, regulator and industry have a good understanding of, and information about, new legal requirements.
- **Effectiveness** Measures are targeted at reducing the risk of online harms, particularly to under 16s. Measures minimise unintended consequences and balance impacts on human rights (particularly children's rights) with the objective of protecting children from risk of exposure to harm.
- **Ease of implementation** Measures are informed by existing or developing technology, utilise current system capability and capacity where possible, and are financially sustainable in the medium and long term.
- **Flexibility** Accommodates future developments in technology, such as the development of new types of platforms and online harms.

41. The criteria align with the policy objectives identified in the problem definition and are weighted differently across the design choices to reflect the distinct purpose and context of each choice. Across the design elements analysed, greater emphasis is placed on the effectiveness criterion to ensure the policy achieves its objective of reducing children's risk of exposure to online harm for children.

42. Some criteria inevitably involve trade-offs. Stakeholders strongly value certainty, clear rules and consistent definitions, but achieving this can reduce flexibility, making it harder for the regime to adapt to rapidly emerging technologies, new platform types, or evolving harms. Conversely, approaches that maximise flexibility and future-proofing can create early uncertainty for platforms and reduce public confidence in how protections will operate in practice. Options that are highly effective at addressing a wide range of harms can also require more complex system changes, which affects ease of implementation. These interactions mean that no single option will perform strongly against all criteria, and each must be assessed on how well it balances immediate protection of under-16s with practical delivery and longer-term adaptability.

43. These design choices also engage several rights, including freedom of expression, peaceful assembly and association under the New Zealand Bill of Rights Act 1990 (NZBORA), and children's rights to participation and access to information under the Children's Convention. Requiring platforms to restrict account-holding for under 16s will limit some opportunities for children to express themselves online, access certain forms of information and participate in peer and community networks. Officials completed a Child Impact Assessment and assessed these limitations as justified and proportionate given the objective of promoting the protection of children from the risk of exposure to online harm, and the presence of meaningful mitigations that reduce the extent of the

rights limitations. Some of these mitigations are listed below and additional mitigations relevant to each design choice are discussed in the analysis that follows.

- **A focus on restricting account-holding only**, which ensures children can still view information and educational content online, but without the personalised feeds or interactions that heighten risk.
- **Continued access to alternative platforms that offer safer forms of connection**, so low-risk, child-appropriate online environments remain accessible.
- **Continued access to help-seeking pathways**, ensuring children can reach services such as Youthline and other verified supports without needing a social-media account.
- **Public education about age assurance**, including how it works and the privacy-preserving criteria that must be met.

What scope will options be considered within?

44. The scope for this impact analysis is intentionally narrow, reflecting Cabinet's December 2025 decision to develop options for Phase One intervention focused on reducing the risk of online harm to children through a minimum age of 16 for social media and other accounts, alongside platform reporting requirements [SOU-25-MIN-0177]. Phase One is not intended to deliver a comprehensive online safety regime; instead, it provides a targeted, response to escalating harms while options for broader online safety reforms are developed for Phase Two.
45. In line with Cabinet's decision, the scope of this analysis excludes broader reforms such as changes to broadcast or video-on-demand regulation, system-wide duties of care, or the creation of a unified content-regulation framework. Only options that can be designed, legislated, and implemented rapidly, and that do not pre-empt Phase Two, are considered feasible within this Phase One assessment.

Constraints shaping the scope of analysis

46. In addition to the Phase One options constraint set by Cabinet, there are several constraints that have materially shaped the scope of feasible options and the depth of analysis:
 - **Time constraints (Cabinet direction):** Phase One was commissioned as a more immediate intervention. Policy development, targeted consultation and option testing were undertaken under compressed timeframes, limiting detailed modelling, wide consultation and full testing of alternative policy interventions.
 - **Cost constraints (Ministerial and Cabinet expectations):** Phase One must be deliverable with minimal additional Crown expenditure and must avoid imposing disproportionate burdensome requirements on platforms and consumers. This limits options that would require significant new infrastructure, bespoke system changes, or extensive ongoing monitoring.
 - **Novelty of minimum-age regimes internationally (evidence constraint):** Few jurisdictions have implemented social media age-thresholds, and available evidence on long-term outcomes, enforcement challenges and system-wide effects is emerging and limited.
 - **Limited NZ-specific data (market constraint):** Many platforms do not routinely share detailed, NZ-specific harm data, making it difficult to quantify local prevalence of

harm, financial impacts or platform-specific risk profiles. The analysis relies on domestic survey data, targeted engagement, and international evidence rather than comprehensive local operational data.

- Rapid technological change (novel policy area constraint): The fast-moving nature of platforms and the technology used by them reduces the confidence in long-term projections and introduces uncertainty into the traditional impact analysis. This includes technology such as AI-driven features (including AI companions) and new age estimation technologies (such as biometrics).

47. International alignment, particularly with Australia: As most platforms operate globally, New Zealand must avoid bespoke technical requirements under Phase One that cannot be practically implemented in a single small market. Alignment with Australia is especially important because New Zealand and Australia share platform markets, many platforms are already adapting to Australia's regime, and alignment reduces compliance costs, supports cross border enforcement, and lowers the risk of New Zealand being deprioritised by platforms.

Other options considered and discounted due to Phase One scope and constraints

48. A range of regulatory and non-regulatory approaches were considered; however, many were discounted because they cannot provide the immediate and operationally feasible response required in Phase One to meaningfully reduce the risk of harm to children under 16. Non-regulatory approaches, in particular, would not sufficiently reduce harm to children online. The current system relies heavily on voluntary action by platforms, which is inconsistently applied and has not been effective in reducing harm. Although most platforms nominally restrict access to users aged 13 and over, these minimum age rules are self-reported- and poorly enforced, and 2025 survey data from NZ on Air found 24 percent of children aged 5 to 14 use social media.²²
49. We also considered managing access to devices through tools such as curfews or education campaigns that support parents to apply parental controls or for children to understand why they should manage their own access. However, we are unaware of any evidence that supports the success of non-regulatory options in isolation of regulatory change, particularly given the scale and immediacy of harms experienced by children.
50. Some platforms and devices offer parental-control tools, but their effectiveness depends on caregiver capability and consistent use, varies significantly across services, and does not prevent children accessing harmful content across multiple devices or accounts. Evidence shows that children under 16 continue to access social media and report harmful online experiences despite parental controls and platform safety features, indicating these measures do not adequately prevent exposure to harm. The Australian Age Assurance Technology Trial found that parental controls are effective only alongside other regulatory measures, not in isolation.²³ The Australian eSafety Commissioner similarly advises that parental controls should form part of a broader set of safeguards and cannot be relied upon to prevent online harm.²⁴ This evidence supports the need for a regulatory intervention as part of Phase One.

²² New Zealand On Air, [Where Are The Audiences - Children's Media Use Report](#), 2025.

²³ Age Check Certification Scheme, [Age Assurance Technology Trial: Part G: Parental Control](#), August 2025.

²⁴ eSafety Commissioner, [Parental controls | How to keep your child safe](#). Accessed on 22/10/2025.

What options are being considered?

Primary intervention options

51. The options analysed differ from the interim RIS to reflect that Cabinet has made the decision to develop a two-step plan to improve children's online safety [recommendation 5 of SOU-25-MIN-0177 refers]. Under Phase One, Cabinet agreed to develop options for a minimum age of 16 for social media and other accounts and imposing child safety reporting requirements on platforms to assess and mitigate the risks of harms to all children under 18 [recommendation 9 of SOU-25-MIN-0177 refers]. This section considers the best primary intervention option to deliver on the first phase of work to reduce the risk of online harm to children. The counterfactual option represents Cabinet's decision that Phase One includes minimum age and reporting requirements for specified online platforms.
52. For the primary intervention, effectiveness and ease of implementation carry the greatest weight, reflecting Phase One's aim to deliver a practical, achievable intervention that can be implemented quickly and reduces the risk of exposure to online harm for under 16s.

Option One – Counterfactual (*preferred*)

53. Cabinet has agreed to develop options for Phase One which would include two key requirements on platforms:
 - A minimum age of 16 to have an account.
 - Reporting on how they are assessing and mitigating risks of harm to children under the age of 18.

Minimum age restrictions

54. Cabinet agreed to develop an approach requiring platforms to take reasonable steps to prevent under-16-year-olds from having accounts and use age assurance methods that align with the Privacy Act 2020 and domestic and international standards.
55. This would require a platform to take reasonable steps to prevent under 16-year-olds from having an account and to use age assurance methods that align with New Zealand's human rights and privacy obligations as well as international standards.
56. A 'reasonable steps' approach would enable a platform to determine how it chooses to prevent under 16s from holding an account provided its methods are effective and exercised in a way that protects a user's dignity and preserves the privacy of their personal information. There would be specific privacy requirements, in addition to the Privacy Act 2020. The regulatory guidance would set out the appropriate standards to align with domestic and international best practice. This is the approach taken by Australia under their Online Safety Amendment (Social Media Minimum Age) Act 2024 which, according to discussions with regulatory stakeholders during targeted consultation, has seen early success.
57. Platforms would be required to provide information on the systems and processes in relation to the age assurance methods, as well as how they are proactively addressing risk of circumvention.

Child safety reporting

58. This would require the platform to have a duty to carry out an accurate assessment of risks to children and report on how it is, or intends to, mitigate the risks. Platforms would

need to complete reporting at least annually but may be required to undertake additional reporting if they make significant changes to their platform.

59. The individual requirements of the reports would be determined through legislation, with decisions still to be made on which elements sit in primary legislation, and which are determined by regulations. Some or all of the detailed reporting requirements could be set in secondary legislation, providing flexibility to update the scope of reporting as evidence and online risks evolve. These requirements would include the list of risks that platforms must report on, how the platform is mitigating those risks, and the methodology used to produce the report. Additionally, before making any significant change to its design or operation, a platform would be required to produce a risk assessment relating to the impacts of the proposed change. It would also be required to nominate an accountable senior manager for the child safety reporting.

Option Two – Allow platforms to create 'teen accounts' for underage users

60. This option would adjust the counterfactual's minimum age requirements to instead enable platforms to create a separate account for people under the age of 16 (also known as 'teen accounts'). Under this option, a platform could choose to invest in creating teen accounts that meet specified requirements for underage users or implement minimum age restrictions if it did not want to invest in the teen account alternative.
61. Many platforms we engaged with did not consider a minimum age restriction would be effective at reducing online harm to children and that there is still online harm in many logged-out platforms. These platforms instead advocated for the use of accounts with reduced accessibility and parental controls as an alternative to minimum age restrictions.
62. Teen accounts limit features for its users (such as only allowing the user to message and tag other users they know), introduces new features to enhance user safety (such as time limiting the user's engagement with the platform) and providing parental controls to give parents increased control and supervision of their children. Teen accounts have already been implemented by several platforms, such as Meta Teen Accounts (for Instagram, Facebook and Messenger) and YouTube has supervised accounts that link to a parental account.

Option Three – Rely on current voluntary reporting by platforms

63. This option would remove the mandatory child safety reporting requirements from the counterfactual and instead rely on current transparency reporting that many platforms already provide. This reporting is not required by any legislation and is provided voluntarily.
64. Typically, most transparency reports produced by platforms include how they are handling user data, responding to government requests and moderating content. In New Zealand, platforms that are signatories to the Aotearoa New Zealand Code of Practice for Online Safety and Harms²⁵ produce annual reports covering their policies, processes and systems for online safety and content moderation and their effectiveness to mitigate risks to users. These reports cover efforts made by the platforms globally but also provide some information and data specific to New Zealand users.

²⁵ This includes Meta (Facebook and Instagram), Google (YouTube), TikTok, Twitch and X (formerly Twitter).

Option Four – Require platforms to have a duty of care to protect children

65. This option would include placing a duty of care on platforms to protect children, in addition to the minimum age and reporting requirements. A digital duty of care would require a platform to proactively remove illegal and harmful content and adopt a 'safety-by-design' approach to their design and systems. Platforms that do not uphold this duty would be subject to penalties.
66. A common theme emerging from targeted consultation on Phase One proposals was that the minimum age and reporting requirements are not enough to adequately protect young people from online harm. Many stakeholders raised that a restriction on social media for under 16s does not directly address the harmful content or harmful behaviours that exist on these platforms, it only delays young people's interaction with them.
67. A duty of care is a legal obligation that is imposed on an individual, requiring adherence to a standard of reasonable care to avoid careless acts that could foreseeably harm others. Duties of care are common in industries where there is a reasonable expectation that a person is responsible for the harm that could result from a person's actions. For example, manufacturers owe a duty of care to consumers of their products; health professionals owe a duty of care to their patients; and directors owe a duty of care to the companies they manage.
68. The UK established a new duty of care for online platforms under its Online Safety Act 2023. This requires platforms to take action against illegal content, or legal content that could be harmful to children where children are likely to access it. Platforms that fail this duty are liable to substantial penalties.
69. In 2024, Australia announced that it was developing a possible digital duty of care regulatory model alongside a minimum age of 16 for access to social media platforms. The proposed duty of care on online services would see the responsibility for managing online harms shift from individuals to the services themselves. Public consultation on a digital duty of care ended on 7 December 2025 and the Australian Government has not made any subsequent announcements about its progress.

How do the options compare to the counterfactual?

	Option One – Counterfactual (preferred)	Option Two – 'Teen accounts'	Option Three – Voluntary reporting	Option Four – Adding a digital duty of care
Certainty	0 Minimum age and reporting requirements are being implemented in multiple jurisdictions, and many platforms are proactively complying.	- It would be more challenging for parents and caregivers to know which platforms are restricted and which allow teen accounts.	0 Certainty would be the same as the counterfactual as voluntary reporting would have no regulatory element.	- A digital duty of care would be novel and is largely untested internationally.
Effectiveness	0 The new requirements reduce the risk of online harm to children by preventing their access to platforms that can expose them to high levels of harm and requiring platforms to report on how they are mitigating risks.	- Evidence has shown that many tools used by teen accounts to keep young people safe are ineffective.	-- Voluntary reporting would be significantly less effective as it provides no consistency in data provided or guarantee that reporting would take place.	++ Inclusion of a digital duty of care would significantly improve online safety outcomes for children.
Ease of implementation	0 The new requirements are immediate measures that can be implemented relatively quickly to reduce the risk of online harm.	- Allowing platforms to choose between teen accounts or age-assurance once deemed in scope would add complexity and create a greater assessment burden for the regulator.	+ Voluntary reporting would provide the regulator with less requirements to implement.	-- Creating an additional duty of care would make implementation of Phase One significantly more complex.
Flexibility	0 The new requirements would apply to the platforms that possess features which contribute to risks of harm to children.	+ Teen accounts would provide more flexibility to platforms regarding how to meet minimum age requirements.	0 Flexibility would be the same as the counterfactual as voluntary reporting would have no regulatory element.	0 A digital duty of care would likely apply to the same platforms as the counterfactual.
Overall assessment	0 The new requirements are intended to be the first step to reducing online harm for children whilst more comprehensive online safety regulations are implemented.	-2 Teen accounts would provide more flexibility for platforms to comply but at the cost of efficacy and ease of implementation.	-1 Voluntary reporting would make it easier to implement Phase One, but the information provided would be significantly less useful.	-1 Including a new digital duty of care in Phase One would greatly reduce the risk of online harm to children but would be much more technically complex to design and implement.

What option is likely to best address the problem, meet the policy objectives, and deliver the highest net benefits?

70. Based on the analysis, the counterfactual of progressing with minimum age and reporting requirements for Phase One is the preferred option. This reflects the weighting applied to the primary intervention, where effectiveness and ease of implementation carry the greatest weight.
71. The counterfactual is expected to reduce exposure to online harm for under-16s by limiting access to account based- features that heighten risk. This includes reducing exposure to harmful content, harmful conduct, risks to privacy and wellbeing, and design features such as infinite scrolling and algorithmic content feeds which can contribute to problematic use. While it would not remove exposure to harm entirely, restricting access to the parts of platform functionality most closely linked to these risks is expected to reduce their frequency and intensity for most under-16s.
72. The preferred option does limit some rights for under-16s, particularly freedom of expression, peaceful assembly and association under the New Zealand Bill of Rights Act 1990 (ss 14, 16–17), and children’s rights to expression, association and access to information under the Children’s Convention (articles 13, 15, 17). The preferred option also supports children’s rights to protection and healthy development under the Children’s Convention (articles 19 and 6) by reducing exposure to harmful online environments. On balance, these limitations are proportionate and justified, as the restriction applies only to account holding and children can still access non-account-based content where available, and use safer platforms for connection.
73. Some children will attempt to bypass a minimum-age restriction, which limits the extent to which the intervention can prevent harm. A certain level of circumvention is expected and cannot be fully prevented. By requiring platforms to monitor and respond to circumvention methods, this would reduce the ability of most under-16s to create or maintain accounts. Some displacement to services with fewer safeguards is likely to occur, this risk is moderated by ensuring that the Phase One scope is flexible enough to capture additional platforms with similar risk profiles. Overall, the intervention is expected to be effective in reducing the risk of online harm for most children.
74. Child safety reporting is an important accountability tool that requires platforms to be transparent about the risks they pose to children and the steps they are taking to reduce them. It would provide the regulator and policy makers with information to help identify trends in online harm and inform further regulatory development, and improve public understanding of platform risks through published reports or summaries, helping people to make informed decisions regarding the platforms they choose to engage with. To mitigate the risk of platforms taking a conservative approach and overreporting on child safety, the regulator would provide guidance to platforms on how to comply with child safety reporting requirements.

Implementing 'teen accounts' would be challenging and have been shown to be less effective

75. Providing platforms with the flexibility to implement teen accounts would make implementing Phase One more complex for the regulator to evaluate whether a platform is meeting legislative requirements. While this option may help to mitigate the risk of young people migrating to less-safe alternative platforms, the efficacy of using teen accounts to reduce online harm for young people is questionable. A 2025 study by Meta

whistleblower, Arturo Béjar, found that two-thirds of safety tools he tested were ineffective and only 17 percent of the tools were working as described by Meta.²⁶

Voluntary reporting would not consistently provide information or guarantee its efficacy

76. Voluntary reporting provided by platforms is inconsistent, as there is no standardised reporting and methodologies can differ greatly between platforms. This limits the value that voluntary reporting may provide to policy makers, regulators and the public. Platforms could also choose to stop their voluntary reporting at any time or choose to withhold specific aspects of their reports (such as the efficacy of their content moderation).

A digital duty of care is better considered under a future phase of online safety

77. For the purposes of implementing Phase One, including a digital duty of care would be too onerous to design and implement. A digital duty of care is a novel concept in New Zealand and remains largely untested globally, aside from early implementation in the UK and EU. Significantly more analysis would be required to create a digital duty of care that works within the New Zealand context, as it would need to account for the substantive case law surrounding duties of care generally. A digital duty of care also needs to be designed in a way that minimises uncertainty, making it clear what obligations apply to platforms and what rights the public can expect. Platforms are still considering what is required to uphold a possible digital duty of care, how it may impact their business models and what changes they may be required to make.

²⁶ Arturo Béjar, et al. [Teen Account, Broken Promises: How Instagram Is Failing To Protect Minors](#), September 2025.

Design options

78. Assuming the preferred primary intervention option is taken (Option One – minimum age and reporting requirements on platforms), this section analyses four separate design questions for how the requirements would operate in practice. The counterfactual for this analysis represents the option that we consider is consistent with our preferred option for primary intervention.

Design question one: Which platforms should be captured by minimum age and reporting requirements?

79. Cabinet has agreed that in addition to prioritising capturing at least the same platforms as Australia for minimum age and reporting requirements, additional platforms should be considered based on what can be done on the platform and the risk of harm to children [recommendation 11 of SOU-25-MIN-0177 refers].
80. Feedback from targeted consultation indicated concern that Australia's narrow, purpose based scope would not capture many of the environments where children spend significant time- and experience harm. Youth, Māori and Pacific groups, and -child rights and mental health- organisations noted that harmful contact, conduct and content frequently occur on platforms excluded under Australia's rules, including gaming environments such as Roblox and Fortnite, hybrid and messaging-based platforms such as Discord, and AI tools that simulate relationships or emotional connection.
81. Platforms valued the clarity of full alignment with Australia but cautioned that harmful experiences are often driven by features rather than platform category, and that a narrow scope could shift under-16s into less-moderated environments. Platforms also emphasised the need for early regulator engagement to avoid the uncertainty experienced during Australia's rapid rollout. As scope decisions need to balance certainty with flexibility to respond to emerging harms, the criteria used to analyse the options under this design question are weighted evenly, while effectiveness and ease of implementation remain key considerations in assessing the options.
82. Scope decisions also affect children's rights to expression, association and access to information (UNCRC articles 13, 15, 17), but these impacts are mitigated because lower-risk communication options such as text, email and some other direct-messaging services, would not be captured under any of the options. This helps preserve children's ability to stay connected with peers and family while focusing minimum-age restrictions on higher-risk environments.

Option One – Counterfactual

83. Under this option, New Zealand would replicate Australia's approach by applying a minimum age of 16 and associated reporting requirements to platforms whose sole or primary purpose is to enable online social interaction between two or more end-users. In practice, this captures the major social-media platforms that provide scrolling feeds, follows or subscriptions, user-generated posts, and public interaction at scale. As in Australia, services whose primary purpose is messaging, calling, online gaming, professional networking, education, health, or school-to-learner or health-provider communication would be excluded. These criteria result in a relatively narrow set of platforms regulated under Phase One.
84. The main advantages of this option are clarity, simplicity and speed. The model has already been designed and implemented in Australia, meaning platforms could reuse

much of the technical infrastructure they have developed for the Australian regime. A New Zealand regulator could draw on Australia's definitions, rules and guidance, reducing early design effort and providing a high degree of certainty for regulated parties.

85. However, Australia's model sits within a broader and more mature online-safety framework that includes mandatory industry codes, stronger children's privacy protections and an independent regulator with extensive statutory powers. New Zealand does not have an equivalent foundation. Replicating Australia's definitional model alone would therefore not reflect New Zealand's context, nor deliver a similar level of protection or system-wide support for reducing risk of harm.

Option Two – Inclusion of social AI companions (preferred)

86. Under this option, the minimum age of 16 and associated reporting requirements would apply to the same social-media platforms captured under Australia's model and would also extend to social AI companions. Social AI companions are AI systems designed to simulate interpersonal or emotional relationships with users through persistent, personalised interactions and companion-style personas. These can include "friend", "mentor", or "romantic" framing, or conversational styles intended to cultivate emotional intimacy. Including social AI companions recognises that children now engage socially not only with other people, but also with AI systems that emulate social connection.
87. Stakeholders consistently identified social AI companions as a high-risk environment for children. Youth, Māori and Pacific groups, child-rights advocates and mental-health experts highlighted the risks of grooming, coercive-control behaviours, emotional dependency, blurred lines between safe and unsafe interactions, and the normalisation of unhealthy relationships. They noted that these systems can replicate, and in some cases intensify, harms observed on social media by creating a sense of trust, intimacy and personalised attention that can be exploited.
88. Platforms providing social AI companions would be required to meet the same requirements as social-media services. To maintain a proportionate, risk-based approach to AI regulation consistent with Cabinet's direction [ECO-24-MIN-0119 refers], this option would require drafting that clearly distinguishes relational AI companions from general-purpose or task-based chatbots. This ensures that commonly used AI services, such as search assistants or productivity tools (such as ChatGPT, Gemini, Siri, Alexa), are not captured, and that children's wider engagement with generative AI for education, health or creative purposes is not restricted.
89. This option also recognises that some AI companion platforms already set an 18+ minimum age internationally; in such cases, a pragmatic regulatory approach can confirm there is no expectation to lower existing minimum ages. Targeting companion style AI avoids the wider complexity of regulating all AI chatbots or gaming, while addressing a clear emerging harm.

Option Three – Inclusion of online multiplayer videogames

90. Under this option, the minimum age of 16 and associated reporting requirements would apply to the platforms captured by Option Two and would also extend to online multiplayer games with communication features. This would include games that enable users to interact through open or private chat, voice communication, friend requests, matchmaking, trading, or user-generated content. These services are not captured under Australia's purpose-based approach because their primary purpose is gaming rather than social interaction.

91. This broader option was included in consultation because of clear evidence of harm in gaming environments, both internationally and within New Zealand, and the high level of gaming use among children. Research and reporting show that children can be exposed to grooming, bullying, inappropriate behaviour via chat, financial exploitation, and harmful user-generated content on gaming platforms. The New Zealand Plays 2025 survey shows that 83% of 5–9-year-olds and 94 percent of 10–14-year-olds play video games.²⁷ This widespread use means many children are routinely exposed to communication features that increase their likelihood of encountering harm.

²⁷ IGEA, [New Zealand Plays 2025](#), 2025.

How do the options compare to the counterfactual?

	Option One – Counterfactual	Option Two – Australia + social AI companions (<i>preferred</i>)	Option Three – Option Two + online multiplayer videogames
Certainty	0 Mirrors Australia's model so platforms already understand the scope. Provides high certainty about which services are captured.	- Certainty is reduced due to the need to define boundaries between companion-style and general-purpose AI.	-- Captures a wide and diverse group of gaming services. Requires feature-level assessments, reducing certainty for both industry and regulators.
Effectiveness	0 Captures major social-media platforms but leaves other significant harm environments unregulated.	++ Captures high harm environments, directly reflecting evidence of emerging harms.	+ Addresses harms in gaming environments, but effectiveness is reduced by definitional challenges and interaction with the classification regime.
Ease of implementation	0 Most straightforward to implement. Platforms can reuse Australian compliance systems and regulators can rely on Australian guidance.	- Requires defining companion-AI boundaries, but once defined, implementation mirrors social-media obligations and complexity remains manageable.	-- Complex to implement because of the large number of additional parties and interaction with the classification regime. Likely to delay commencement.
Flexibility	0 Least flexible. Scope is tied closely to Australia's purpose-based definition and does not adapt easily to emerging harms.	+ More flexible. Companion-AI criteria can be refined over time to maintain a proportionate, risk-based scope without capturing broader AI tools.	++ The most flexible option as it responds to the widest range of harms across different platform types. Allows the regulatory framework to adapt to diverse environments where children interact and experience risk.
Overall assessment	0 Provides clarity and relative ease of implementation but the narrow scope limits ability to address the range of harms identified.	+1 Applies a targeted, risk-based scope that captures emerging harms while remaining manageable for Phase One implementation.	-1 Captures more harm environments but introduces disproportionate complexity and delay, making it unsuitable for Phase One.

What option is likely to best address the problem, meet the policy objectives, and deliver the highest net benefits?

92. Option One is the option likely to best address the problem and meet the policy objective. This approach extends the counterfactual in a focused way by applying the minimum age to both social-media platforms and social AI companions. Evidence from research and targeted consultation shows that children encounter harmful content, contact and conduct in these environments, driven by features such as chat, private messaging, user-generated content and personalised interaction.
93. These features appear across multiple services including those not captured by a purpose-based social-media definition. Expanding the counterfactual to include social AI companions fills a clear regulatory gap and is expected to reduce harm while keeping Phase One implementable. It also provides greater flexibility than a narrow purpose-based scope by allowing emerging platforms with similar features that present risks to children's safety to be brought into scope, while still maintaining a reasonable degree of certainty for platforms about how boundaries will be drawn. Flexibility is further supported by the ability to refine which classes of platform are in or out of scope through regulations, allowing the framework to evolve as platform types and risk profiles evolve.

Full alignment with Australia limits flexibility to respond to emerging harms

94. While aligning fully with Australia provides clarity and ease of implementation, a purpose-based definition would restrict Phase One to traditional social-media platforms, even though harmful features are found across a much wider range of services, including hybrid platforms, gaming environments and AI companions.
95. In Australia, the minimum-age requirement operates within a broader and more mature online-safety framework, that includes mandatory industry codes, strengthened children's privacy protections, and an independent regulator with extensive enforcement powers. New Zealand does not yet have these system-level supports and replicating the Australian social media minimum age model alone would deliver limited benefits and leave key harm environments unaddressed. It would also risk New Zealand falling further behind Australia, which is already expanding its online-safety model, including the new industry codes for 18+ AI chatbot content that are now in effect.

Regulating gaming environments is not feasible under Phase One constraints

96. Extending the minimum age to online multiplayer games would capture additional harm environments but introduces substantial implementation challenges. Applying a statutory minimum age to games risks overlapping with the existing classification regime, which does not require online games to carry mandatory age labels and could create inconsistency across regulatory settings. It would also bring a wide range of additional providers into scope, including console and game storefront providers such as Microsoft (Xbox) and Steam, significantly increasing complexity and potentially delaying commencement. These factors make gaming-related measures better suited to Phase Two, when broader system settings and regulatory foundations will be more developed.

Design question two: How should platforms comply with minimum age requirements?

97. Cabinet agreed to develop an approach requiring platforms to take reasonable steps to prevent under-16-year-olds from having account and use age assurance methods that align with the Privacy Act 2020 and domestic and international standards [recommendation 12 of SOU-25-MIN-0177 refers].
98. Feedback from experts on age assurance technologies highlighted that there is no 'one size fits all' approach to determining a user's age or age-range, and that regardless of the method chosen, there would be trade-offs between accuracy, privacy and data minimisation. Many youth, Māori and disabled peoples' stakeholders focussed on the need to provide multiple ways for a user to determine their age and that the methods must protect the user's privacy. The criteria for analysing this design question are weighted evenly to balance certainty for users and platforms with the flexibility needed to accommodate different and emerging age assurance methods, alongside considerations of effectiveness and ease of implementation.
99. Requiring platforms to implement a minimum age engages children's privacy rights under the Children's Convention (art 16), as age assurance technology requires some form of data handling. Under both options, additional privacy protective requirements would apply, beyond those in the Privacy Act 2020, including limiting the use of data for age assurance purposes only. Education about age assurance can also help children and families understand what information is being used and why. These measures ensure that any limitation on privacy is proportionate to the objective of reducing harm and support children's rights to protection and healthy development (articles 19 and 6).

Option One – Counterfactual (*preferred*)

100. Under the counterfactual, platforms would be required to take a high-level outcomes and performance-based approach to implementing 'reasonable steps' to prevent people under the age of 16 from holding an account. This is a permissive approach that would give platforms the freedom to choose how they determine a user is over the age of 16, provided the methods were reasonably effective and upheld privacy standards.
101. The regulator would provide guidance to platforms regarding how they may comply with minimum age requirements. In the event of non-compliance (existing or suspected), it would be empowered to respond and create rules stipulating how a platform must or must not meet minimum age requirements. These rules would be secondary legislation and could include matters such as requiring a platform to use certain age assurance measures (such as a method accredited under the Digital Identity Services Trust Framework²⁸ or prohibit age assurance methods that are known to be ineffective such as self-attestation).
102. The counterfactual would also prohibit platforms from relying solely on formal ID or other accredited digital credentials. This prohibition recognises that not everyone has these forms of ID or, if they do, may not feel comfortable sharing it with a platform.
103. The counterfactual follows the approach taken by Australia in its Online Safety Act which has seen some early success with approximately 4.7 million accounts being removed

²⁸ The Digital Identity Services Trust Framework is the legal framework for digital identity services in New Zealand. The framework includes rules and regulations for how digital identity services that are accredited should work, protecting information and privacy.

from platforms.²⁹ The proposal for a regulator to respond to non-compliance with specific direction on age assurance methods is an additional feature which we understand does not exist in eSafety's social media minimum age powers.

Option Two – Prescriptive requirements outlining how a platform must comply with minimum age requirements

104. This option would have the minimum age requirements and 'reasonable steps' prescribed in legislation, as opposed to creating a high-level duty under the counterfactual.
105. Specifically, this option would state what age assurance methods a platform would have to use to determine whether a user is over the age of 16, the approved technologies that a platform would have to use and the approved age documents that users would have to use (such as a passport or driver's licence). This option could include requiring a platform to only use age verification methods that have been accredited under the Digital Identity Services Trust Framework (monitored by a separate regulator – The Trust Framework Authority), whilst still requiring platforms to offer other age inference and estimation methods to ensure no reliance on ID provision alone.
106. This option follows the approach taken by New Zealand's Sale and Supply of Alcohol Act 2012, which states that verification of a person's age must come from approved evidence of age document or through approved evidence of age system.

²⁹ eSafety Commissioner [Platforms restrict access to 4.7 million under-16 accounts across Australia](#), January 2026.

How do the options compare to the counterfactual?

	Option One – Counterfactual (<i>preferred</i>)	Option Two – Prescriptive requirements
Certainty	0 Platforms would know that there is an expectation that they implement minimum age requirements and they must meet some objective measures.	0 + Prescriptive requirements would better inform platforms of how they have to comply with the legislation and inform the public of how they can demonstrate they are over 16.
Effectiveness	0 Efficacy can depend on the age assurance method and technology used but must still be reasonably effective at preventing under-16s from holding an account.	0 + Prescriptive requirements would ensure that a high level of efficacy for age assurance is maintained across all platforms operating in New Zealand.
Ease of implementation	0 Most platforms have already taken reasonable steps to comply with minimum age requirements in Australia that could be applied in New Zealand.	0 - The regulator would likely need to devote additional resources to ensure that platforms were meeting specific minimum age requirements in the legislation.
Flexibility	0 A high-level requirement would provide platforms with a lot of flexibility to determine how they implement minimum age requirements. This also benefits the user as they are provided a choice in how they would like to complete age assurance.	0 -- Prescriptive requirements would not be flexible or easily updated if new age assurance technologies are developed. It also does not allow platforms to have as flexible an approach to compliance in New Zealand, and users may have less choice in how they demonstrate they are over 16.
Overall assessment	0 A high-level requirement would maintain efficacy whilst staying robust to technological changes over time and enable platforms to apply age assurance methods consistently across jurisdictions.	0 -1 Prescriptive requirements would provide greater certainty and consistency but would become outdated quickly given the pace of technological change.

What option is likely to best address the problem, meet the policy objectives, and deliver the highest net benefits?

107. Based on the analysis, the counterfactual remains the preferred option. Adopting a high-level approach to minimum age requirements that focusses on outcomes and performance would make it clear to platforms that they have a duty to prevent people under the age of 16 from holding an account, setting clear privacy-preserving requirements and giving platforms the flexibility to choose how they meet the requirements. The counterfactual performs well across the criteria, with its flexibility for platforms and greater user choice making it preferable to a prescriptive approach.
108. Flexibility is essential for the minimum age requirements to be implemented effectively. Many jurisdictions are going through the process of implementing minimum age requirements and platforms will be investing significantly in age assurance technologies to meet these new requirements globally. A high-level approach would provide enough flexibility for platforms to determine the best way to comply with New Zealand's legislation, thereby making it likelier that platforms continue to operate in the New Zealand market. As well, a high-level approach would be resilient to technological changes and ensures that the minimum age requirements remain relevant and effective over time.
109. The counterfactual would prevent a 'race to the bottom' style of compliance by placing some objective requirements on the platforms (such as complying with privacy legislation and not relying solely on one form of age assurance) that would guide platforms towards adopting effective age assurance methods. The regulator would also be able to provide guidance to assist compliance with the high-level requirements, thereby ensuring efficacy is maintained.
110. The prescriptive approach would provide more clarity to platforms and ensure a higher level of consistency for minimum age requirements across platforms. However, this approach is less flexible and could quickly become outdated. It also limits user and platform choice by requiring specific age assurance methods to be used, regardless of whether they suit all users/services. For users, this could mean being required to use age assurance methods they are not comfortable with, creating barriers to participation. For platforms, this could require investment in bespoke age assurance solutions that are not easily scalable, and this would be a barrier to them operating in the New Zealand market. While accredited services under the Digital Identity Services Trust Framework are expected to grow over time, availability remains limited in the short term, which could delay implementation and increase compliance burden.
111. Additionally, a prescriptive approach would require the regulator to have more in-depth monitoring of platforms' age assurance methods to ensure they were meeting specific requirements. This is much more operationally burdensome, time consuming and costly, compared to taking an outcome and performance-based approach under the counterfactual.

Design question three: What should platforms be required to report on?

112. Cabinet agreed that platforms should be required to report on the risks of harm to all New Zealand users under 18, including how they assess and mitigate these risks [SOU-25-SUB-0177 refers]. This reporting will help reduce harm for children and young people, support the regulator to build an evidence base for Phase Two, and establish early expectations of transparency and accountability.
113. During consultation, platforms generally preferred that reporting be provided privately to the regulator, noting that detailed assessments can contain sensitive operational information and that producing New Zealand-specific data may require building new reporting systems. In contrast, civil-society groups, researchers, mental-health experts and youth organisations expressed strong support for greater transparency, arguing that some form of public reporting or regulator-issued summaries are essential for independent scrutiny, accountability and building an accurate national understanding of the risks children face online.
114. We considered two options for establishing Phase One reporting requirements, informed by New Zealand's existing voluntary reporting and international models. Analysis of the options under this question places the greatest weight on effectiveness and flexibility, as reporting requirements need to drive meaningful improvements in child safety while remaining able to evolve as new risks emerge.

Option One – Counterfactual (*preferred*)

115. Under this option, legislation would set out a comprehensive list of risks for platforms to assess and mitigate, supported by a statutory duty to complete accurate children's risk assessments and mitigation plans. Platforms would be required to:
- keep assessments up to date and provide them at least annually;
116. complete a child risk assessment before making any significant change to platform design or operation; and
117. report on methodology, instances of harm, and usage and engagement data for New Zealand child users (including logged in users aged 16+ and risks to under 18s in logged out states).
- nominate an individual senior manager responsible for the child safety reports.
118. This option responds directly to findings from the Education and Workforce Committee's interim inquiry, which highlighted limited visibility of how platform systems and design choices expose young people to harm, insufficient transparency around algorithmic processes, and a lack of consistent reporting on how risks to children are identified and mitigated.
119. A comprehensive reporting duty enables the regulator to obtain the full set of information needed to assess these risks, rather than relying on narrow content-harm categories, and establishes clearer expectations that platforms must actively identify and mitigate a range of harms affecting children. It is the start of safety by design requirements for platforms, requiring them to demonstrate what they are doing to be safer for children if they choose to return to their platforms from the age of 16, as well as safer for children who are able to use them (either through an account 16+ or in a logged out state).
120. However, embedding a list of harms in legislation reduces flexibility to update reporting expectations as new risks emerge. It also places higher up-front compliance burdens on platforms, particularly those that do not currently report in New Zealand or do not hold

New Zealand-specific harm data, as they would need to build new internal systems and processes to meet the statutory reporting obligations.

Option Two – Short list of key platform harms aligned with existing voluntary reporting

121. Under this option, legislation would require platforms to report on a short list of key harms aligned with the voluntary Aotearoa New Zealand Code of Practice for Online Safety and Harms. The Code requires reporting on bullying, child-safety disinformation, harassment, hate speech and misinformation, and major signatories, including Google, Meta, TikTok, Twitch and X, already publish annual transparency reports containing information on New Zealand user reports. This approach builds on existing reporting systems and would keep compliance burdens relatively low.
122. However, limiting reporting to a narrow set of content-related harms would not provide the regulator with insight into the wider range of risks children face online, including risks arising from platform design, recommendation systems and interaction features. These existing voluntary categories are also not well-suited to capturing risks from emerging technologies such as AI companions. This narrower approach also offers less flexibility to expand reporting as new harms emerge or as evidence evolves. For these reasons, while this option is simpler for platforms and draws on familiar reporting categories, it would not provide the breadth of information needed to support Phase One compliance decision-making or inform Phase Two options.

How do the options compare to the counterfactual?

	Option One – Counterfactual (<i>preferred</i>)	Option Two – Short list of key platform harms aligned with existing voluntary reporting
Certainty	0 Provides a clear and comprehensive statutory baseline, giving the regulator and platforms certainty about what must be reported.	0 Uses familiar, well-understood categories from existing voluntary reporting, giving similar predictability about what must be reported.
Effectiveness	0 Enables visibility across a wide range of risks (content, contact, conduct, design/algorithmic factors), supporting stronger oversight of harms.	-- Focuses on content-only categories, limiting insight into risks arising from platform design, recommendation systems and interaction features.
Ease of implementation	0 Higher initial burden, especially for platforms without NZ-specific data who would need to build new reporting systems/processes.	+ Easier for platforms to comply with, as it draws on existing voluntary reporting categories and data systems many platforms already use.
Flexibility	0 Some or all reporting detail may be set in regulations, providing flexibility to update reporting requirements and add new harm categories as risks evolve.	-- Statutory scope focused only on content-type harms is less flexible. Even if the harms list sat in regulations, it excludes risks arising from platform design and interaction features. Cannot readily adapt to emerging harm types without legislative change.
Overall assessment	0 More onerous for platforms but provides the breadth of information needed for Phase One oversight and for building an evidence base for Phase Two.	-3 Simpler for platforms but too narrow to support effective Phase One decision-making or to inform future regulatory design.

What option is likely to best address the problem, meet the policy objectives, and deliver the highest net benefits?

123. A comprehensive reporting duty (Option One) is the option most likely to meet the policy objectives. It gives the regulator visibility across the wider range of risks children face online, including those arising from platform design, algorithms, contact features and interaction patterns. This breadth is important for Phase One, as it enables the regulator to understand how harms occur and how platforms are mitigating them, and it provides a stronger foundation for informing future decisions on Phase Two.
124. Larger global platforms already engaged in international reporting are better positioned to comply, while smaller or emerging services, including those that are not signatories to the voluntary Code, such as Reddit and many AI companion providers, may need a longer lead-in time to build the systems required to produce New Zealand-specific reporting. These differences reflect variations in scale, data infrastructure, and existing transparency processes across platforms.
125. Option one is unlikely to affect competition, as major platforms already comply with similar requirements in other jurisdictions. Compliance costs for platforms are expected to fall overtime as reporting systems stabilise and platforms adapt. The assessment assumes clear regulator guidance and alignment with existing reporting frameworks, where possible, to reduce duplication and uncertainty.
126. Non-monetised benefits, such as improved visibility of harms, algorithmic risks and mitigation practices, are assessed as high, as these insights underpin effective oversight and policy development. Non-monetised costs, including the administrative effort of building new reporting systems, are moderate and concentrated among a small number of providers. Risks relate mainly to providers' ability to generate New Zealand-specific data and to managing sensitive operational information in any published summaries; both can be addressed through regulator guidance and regulator-controlled publication.
127. The reporting duty also supports children's rights to protection and healthy development under the Children's Convention (art 19). Under both approaches, the mitigations that platforms implement in response to their child-safety assessments should make platforms safer for children and, in many cases, for all users. However, requiring platforms to assess risks against a more flexible and comprehensive list of harms is likely to lead to stronger and wider-reaching mitigations that improve children's safety online. The greater efficacy and flexibility of a comprehensive reporting duty mean it better meets the criteria and is the option most likely to achieve Phase One's objectives.
128. Flexibility is further supported by the ability to update reporting requirements through regulations, including the specific harms platforms must assess and report on, allowing the regime to adapt as evidence and online risks evolve. Overall, while more resource intensive initially, this option delivers greater long-term value by ensuring Phase Two options are informed by a reliable understanding of the risks facing New Zealand children.

Design question four: How should the regulator penalise serious instances of non-compliance?

129. Cabinet agreed to develop options for a new interim online safety regulator with monitoring and enforcement powers [recommendation 13 of SOU-25-MIN-0177 refers]. As part of its analysis, the Cabinet paper considered providing the regulator with a range of powers to enforce compliance with the legislation. These enforcement tools included significant financial penalties and business disruption tools [paragraphs 41 and 42 of SOU-25-SUB-0177 refers].
130. Feedback from stakeholders representing human rights groups, privacy groups and oversight bodies emphasised that strong regulatory powers are essential for Phase One. Given the need for enforcement tools that can effectively deter serious non-compliance, the effectiveness criterion is weighted most heavily in the analysis of this design question.
131. Other jurisdictions with more comprehensive online safety regimes employ a wide range of enforcement tools to dissuade and penalise non-compliance. At the less serious end, these include administrative actions, infringement notices, enforceable undertakings and injunctions, and at the more serious end these include financial penalties and business disruption tools. For example, Australia has financial penalties of up to NZD\$58.9 million for breaches of their minimum age restrictions, the UK can fine platform's up to £18 million or 10 percent of their global annual revenue, the European Union can fine platforms up to 6 percent of their global turnover and Singapore can require content to be taken down or mandate platforms put in place stricter safeguards (such as parental controls). Business disruption tools are used by online safety regulators in Australia and the UK, and they are recognised as essential tools to reflect the challenges of enforcing rules on platforms with no, or limited, physical presence in a country.

Option One – Counterfactual (*preferred*)

132. Under the counterfactual, the regulator would be empowered to apply for significant financial penalties and business disruption tools in response to serious and/or repeated non-compliance with the legislation.
133. Civil pecuniary penalties are a common enforcement tool used by many regulators both in New Zealand and internationally. While the method of calculating the financial penalties is not being assessed, it is expected that financial penalties under Phase One would need to be very high. Phase One would have platforms within scope with high revenues; therefore, the financial penalties must be proportionate to the platform if they are to be effectively dissuasive and punitive.
134. Business disruption tools are less common enforcement powers. These orders would be used only in the most serious instances of non-compliance and would require judicial authorisation to use. These tools would include:
 135. **Service disruption orders** that require businesses offering ancillary services (such as facilitating financial transfers or enabling advertisements to be shown on the platform) to withdraw their services from a platform.
 136. **Access restriction orders** that require businesses (such as internet service providers or app stores) to stop providing access to the platform, or require the platform to prevent New Zealand users from accessing it or creating new accounts.
137. New Zealand has limited precedent for enforcement powers equivalent to the proposed business disruption tools. The closest parallel to a service disruption order is in the

Commerce Act, which enables the Commerce Commission to seek injunctions preventing a business from knowingly assisting another to contravene the law, although this is narrower and does not extend to directing ancillary providers to withdraw support. The closest comparison for an access-restriction order is the take-down notices under the FVPCA which requires hosts of online content to remove objectionable content. Electronic systems to prevent public access to online platforms (web filter) were considered under the proposed amendment that introduced take-down notices, but these tools were removed during the select committee process due to concerns around the lack of detail and parliamentary oversight, the impact they could have on freedom of expression and that a filter may be an unworkable solution.³⁰

138. The proposed business disruption tools would be novel in the New Zealand context, while aligning with regulatory approaches used internationally. These tools are provided in recognition that there may be instances where it is inappropriate for businesses to continue engaging with a seriously non-compliant platform that is putting under 16-year-olds at risk of harm or for users to continue to have access to a platform.

Option Two – Rely solely on financial penalties

139. This option would reduce the tools available to the regulator for punishing serious and/or repeated non-compliance to solely financial penalties.
140. Financial penalties are a well-established and effective method for enforcing legislation. They are easily understood with existing systems in place to facilitate their operation. Comparatively, business disruption tools are highly invasive enforcement powers that are relatively novel and largely untested internationally.
141. Both in Australia and the UK, business disruption tools have been used very sparingly and their long-term efficacy and impacts are unknown. It may be more pragmatic to rely solely on financial penalties as part of Phase One until the effects of business disruption tools have been tried and tested in other jurisdictions.

³⁰ Governance and Administration Committee, [Films, Videos, and Publications Classification \(Urgent Interim Classification of Publications and Prevention of Online Harm\) Amendment Bill](#), 13 September 2021.

How do the options compare to the counterfactual?

	Option One – Counterfactual (<i>preferred</i>)	Option Two – Financial penalties only
Certainty	0 All enforcement tools would be clearly described in the legislation.	0 There would likely be no change in certainty between requiring financial penalties and business disruption tools <i>and</i> only requiring financial penalties.
Effectiveness	0 Business disruption tools and financial penalties provide the regulator with strong enforcement powers to dissuade and penalise non-compliance with the legislation.	-- Without business disruption tools, there is a significant risk that harmful platforms continue to operate and are accessible by New Zealand children.
Ease of implementation	0 Financial penalties are widely used by New Zealand enforcement agencies, but business disruption tools are novel and would require additional processes to guide their usage.	+ Requiring only financial penalties would mean the regulator has one less enforcement tool to develop internal processes and policies for. The regulator would not get tied up in legal processes associated with the business disruption orders.
Flexibility	0 Providing the regulator with a wide range of enforcement powers for significant offending means it can effectively respond on a case-by-case basis.	- Not having access to business disruption tools means the regulator has less capability to respond to serious non-compliance that may arise under Phase One.
Overall assessment	0 Business disruption tools are essential to dissuade and punish the most serious instances of non-compliance where it would be inappropriate for a platform to remain accessible in New Zealand.	-2 Relying solely on financial penalties to dissuade and punish serious non-compliance would leave the regulator without the tools it needs to effectively enforce Phase One.

What option is likely to best address the problem, meet the policy objectives, and deliver the highest net benefits?

142. Based on the analysis above, the counterfactual remains the preferred option. Business disruption tools are necessary as a last resort to prevent online harm to children (particularly under 16s) from the most non-compliant platforms that have a significantly detrimental impact on children, with no intention of changing their platform or making it inaccessible.
143. Business disruption tools are justified as they would only be used in the most serious instances of non-compliance. The risk with relying solely on financial penalties is that it does nothing to prevent a platform from still being accessed by New Zealand children. Given the need for effective deterrence, financial penalties alone are insufficient to address the most serious non-compliance.
144. Given the cross-border nature of many platforms, there is no guarantee that a non-compliant platform could be compelled to pay a financial penalty. This would limit the regulator's authority and ability to effectively enforce compliance. Business disruption tools overcome this barrier because they can operate through domestic service providers, ensuring the regulator can act even when penalties cannot be collected. This option is therefore recommended over financial penalties on their own, as it is more effective at addressing serious non-compliance.
145. Possible unintended impacts of business disruption tools would be mitigated by requiring the regulator to seek judicial authorisation and placing requirements that the regulator would need to meet. Judicial authorisation would provide the platform with an opportunity to argue why it should not be subject to business disruption measures and potentially appeal a decision it did not agree with. The regulator would be required to show that it had tried other methods of enforcement that had been ineffective before it could apply through the courts for business disruption tools.
146. Requiring judicial authorisation would also reduce the possible impact on businesses being prevented from engaging with a platform by providing a clear signal that a platform is seriously contravening New Zealand legislation. If an application for business disruption tools is successful, this would clearly indicate that a business should not be dealing with the platform that is increasing the risk of children (including under 16s) being exposed to a risk of harm.
147. Digital infrastructure experts highlighted the complexity of filtering and blocking platforms, but that this could be managed by providing clear requirements in the legislation regarding their use and the regulator providing guidance regarding how it would exercise business disruption powers.

Is the Minister's preferred option in the Cabinet paper the same as the agency's preferred option in the RIS?

148. Yes

What are the marginal costs and benefits of the preferred option in the Cabinet paper?

Affected groups	Comment	Impact	Evidence Certainty
Additional costs of the preferred option compared to taking no action			
Platforms	Platforms will incur a one-off setup and ongoing compliance costs. - One-off setup costs: - Age-assurance system build/integration. - NZ specific reporting infrastructure. - Risk-assessment methodology and governance. - Under-16 account deactivation and user notifications. - Data-handling policy for age assurance processes. - Ongoing compliance costs: - Annual children's risk assessment and report production - Operate and maintain age assurance systems. - Routine collaboration with the regulator. - Maintaining information on age assurance systems and processes. Setup costs are modest for large platforms already operating similar systems overseas, and higher for smaller or emerging services, including those not with a minimum age of 16 requirement (incl. AI social companion platforms). Ongoing compliance costs are expected to stay relatively consistent overtime. Based on targeted consultation with platforms, international regulators, and advice from age-assurance experts.	Medium – High	Low – Medium
Regulator	The regulator will incur one-off setup and ongoing compliance costs. - One-off setup costs: - Developing guidance and relationships with platforms. - Building on existing technical expertise and analytics capability. - Designing public campaigns and resources (in consultation with children and whānau). - Establishing evaluation and research processes. - Establishing independent audit processes. - Setting up systems for assessing and publishing summaries, or amended versions, of reports to support transparency. - Ongoing monitoring/compliance costs: - Proactive monitoring of platforms. - Investigations and enforcement activity. - Ongoing communications with platforms. - Delivering public-education activity and updating resources. - Conducting evaluation and research.	Medium	Medium

Affected groups	Comment	Impact	Evidence Certainty
	• Undertaking independent audits. • Ongoing assessment and publication of summaries, or amended versions, of reports. [BUDGET SENSITIVE] Figures subject to Budget 2026 decisions: • \$6.000 million in 2026/27. • \$7.750 million in 2028/29. • \$8.500 million per annum from 2029/30 onwards. These costs were developed using DIA's regulatory function cost model for FTEs, seniority, overheads and timing, based on the assumption that the Phase One regulator would operate as an expanded version of DIA's Digital Safety Group. Final values depend on the enacted powers, audit scope, legal activity, and the design of education and evaluation components.		
Young people	• Under 16s: Loss of access to accounts on regulated platforms and possible displacement to unregulated or less safe services. Noting ability to access to regulated platform in a logged-out state remains for under 16s. • Possible mental and emotional costs from reduced social connection or participation cannot be monetised or reliably quantified. These impacts are recognised qualitatively only. • 16-17s: Friction and potential for incorrect age estimates requiring appeal. Some platforms consider the impact on young people of removing access to accounts as high, noting that doing so would mean children lose access to safeguards built into accounts. These safeguards are not always available, are not applied consistently where they do exist, and many parents are unaware of how to access them. Impacts related to reduced social connection would be mitigated by ensuring that online spaces for young people to interact with others remain accessible. Cost assessment was based on feedback from targeted consultation. Australian experience is emerging but not yet conclusive.	Low – Medium	Low
Wider public	• Friction largely experienced when creating a new account on regulated platform (need for either verification ID), or estimation (facial scan). However, the expectation is that little to no friction should be experienced by users with existing accounts as most platforms can rely on age inference to determine their age without requiring additional checks. • Risk of privacy breach if platforms fail to meet legislative standards for age-assurance processes. This could negatively impact public trust. • Medium privacy/trust cost where assurance messaging is unclear. • 6(a)	Low - Medium	Medium

Affected groups	Comment	Impact	Evidence Certainty
	6(a) Phase One would include steps to mitigate privacy risks, placing stronger privacy requirements on platforms in addition to the Privacy Act 2020. Assessment informed by experience in Australia. Impact dependent on age assurance implementation by platforms & public communications.		
Total monetised costs	A reliable total cannot be calculated now given data gaps and double-counting risks. We therefore present non-monetised impacts and will monetise ranges once Phase One data is available.	N/A	N/A
Non-monetised costs		Medium	Low – Medium
Additional benefits of the preferred option compared to taking no action			
Platforms	- Alignment with emerging international regimes (e.g. Australia, UK, EU) reduces duplication of reporting and age assurance requirements. - Creates opportunities to innovate and develop/adopt safety technologies that improve platform experience for children and, in some cases, general users. - Ability to demonstrate social responsibility through transparent reporting and risk mitigation practices. - Reduces the risk of their services exposing children to harms. Benefits inferred based on international regulatory experience and feedback from platforms.	Low – Medium	Low – Medium
Regulators	- Proposals will provide clarity to New Zealand's online safety regulatory framework and provide a first step in closing existing regulatory gaps. - Improved visibility of platform practices and harms affecting New Zealand children. - Strong evidence base to inform enforcement, prioritisation and Phase Two regulatory design. - Enhanced capacity to identify system gaps and emerging risks early. - Clearer ability to coordinate across government and with international online safety regulators. Benefits grounded in reporting requirements and experience of comparable international regulators.	High	Medium
Young people	Key Benefit: <i>Reduced exposure to harmful content, contact, conduct and algorithm-driven risks on regulated platforms. Particular benefit for children from marginalised groups who face disproportionately higher risks of online harm.</i> Safer baseline experience for 16–17-year-olds under strengthened oversight and platform mitigations.	High	Medium

Affected groups	Comment	Impact	Evidence Certainty
	• Potential shift toward safer alternative activities or platforms. Children will have more time back for developmental activities such as play and sleep. • Downstream wellbeing benefits may emerge over time; these will not be quantified in Phase One. Long-term wellbeing effects will be measured through evaluation. • Establishes a clearer social norm that under-16s are not expected to be on social media or other platforms, reducing pressure and “fear of missing out” for children who may otherwise feel they must participate to maintain social connection. Assessment of benefits supported by evidence on the prevalence and impacts of online harm, including qualitative accounts from the EWC interim inquiry and international research.		
Parents and caregivers	• Greater confidence that platforms accessed by their children are operating under minimum age requirements and safety reporting obligations. • Clearer information about platform risks. • Indirect benefit through reduced exposure of children to harmful interactions online. • Changing norms make it easier for parents to set boundaries and have conversations about online activity and safety, as not being on social media and other platforms under 16 becomes standard rather than exceptional. Assessment of benefits supported by parent feedback that clearer platform rules and age expectations make online-safety conversations easier.	Medium	Medium
Wider government	• Stronger child-safety reporting and improved visibility of platform behaviour will enhance the evidence base available to Government agencies to inform initiatives in education, wellbeing, digital safety, scams, and youth mental health. This supports better targeting of education, health and wellbeing investments, including options for Phase Two online-safety programmes. • Reduced system pressure where online harms are mitigated upstream (e.g., fewer cases escalating into mental-health services). • Aligns with and complements existing Government investments aimed at strengthening children’s wellbeing and positive offline experiences, including: ○ Implementation of the proposed school curriculum, which embeds strengthened online-safety content. ○ Sport NZ’s expansion of the Active Me – Kia Tū Hardship Fund, supporting young people facing financial barriers to participate in sport and recreation.³¹	High	Medium – High

³¹ [More funding means more sport opportunities for children | Beehive.govt.nz](#). October 2025.

Affected groups	Comment	Impact	Evidence Certainty
	○ Netsafe, Classification Office, and DIA education programmes designed to reduce online harm and build digital literacy.³² ○ Supports school-based efforts to reduce digital distractions and promote wellbeing, complementing policies such as ‘Phones Away for the Day’.³³ ○ The Digital Identity Services Trust Framework and Government digital-identity work, which can support privacy-preserving approaches to age assurance (e.g., enabling verified credentials such as a ‘driver licence in your phone’).³⁴ Based on consultation with Government agencies (including the Ministry of Education, Classification Office, Netsafe and DIA), and review of relevant Government initiatives aimed at reducing online harm, supporting digital literacy and improving youth wellbeing.		
Total monetised benefits	A reliable total cannot be calculated now given data gaps and double-counting risks. We therefore present non-monetised impacts and will monetise ranges once Phase One data is available.	N/A	N/A
Non-monetised benefits		High	Medium

³² [Online Safety Resources | Netsafe, Parents & whānau | Classification Office; Keep It Real Online | DIA](#)

³³ Ministry of Education, [Phones Away for the Day](#), August 2024.

³⁴ Department of Internal Affairs, [Trust Framework for Digital Identity Legislation - dia.govt.nz](#). Accessed Feb 2026.

Section 3: Delivering an option

How will the proposal be implemented?

149. Initial implementation of the Phase One legislation would be shared between the DIA policy resource and the Phase One regulator, which is likely the secretary supported by the Digital Safety Group within DIA. DIA would be responsible for developing any regulations needed to bring the legislation into effect, while the regulatory function would separately develop guidance for platforms on their new obligations and provide education to children, parents/caregivers, and the public on the minimum age requirements. The Digital Safety Group would be responsible for enforcing the legislation during Phase One.

150. 9(2)(f)(iv)



151. The Phase One legislation would come into effect six months after it receives Royal Assent. A delayed commencement date provides DIA and the regulator with the opportunity to pass regulations, provide guidance and deliver education. [BUDGET SENSITIVE] Digital Safety preparation work could begin from July 2026, pending Budget 2026 funding to support a 12-month overall implementation period.

152. This also provides platforms with an opportunity to bring their services into line with the new requirements, gives New Zealand children who may be using platforms an opportunity to grow comfortable with having their access restricted, gives parents and caregivers the opportunity to support children in their care before the requirements take effect and gives the New Zealand public an opportunity to understand how the new requirements may affect them.

153. The policy group within DIA would be required to support the passing of regulations to bring the Phase One legislation into force. Although the precise details of Phase One are still being determined, it is expected that DIA could potentially develop regulations that:

- specify classes of platform that are in or out of scope of the Phase One requirements (such as platforms predominantly used for peer-to-peer messaging, platforms for education purposes and platforms for health services)
- provide detail on the risks of harm that platforms would be required to collect information on for their child safety reporting. This could include stating the specific types of harmful content and/or harmful behaviour a platform would need to report on (such as cyberbullying, unwanted contact from adults and spreading disinformation).

154. The Digital Safety Group would develop guidance for platforms, and this could begin following introduction given existing resourcing and hiring capability. This guidance could include helping platforms determine whether they are in scope of the Phase One

requirements, outlining how a platform could comply with the high-level requirements for a minimum age restriction and how a platform could produce child safety reporting.

155. Alongside guidance, the regulator would develop and deliver education on the rationale of minimum age restrictions, provide guidance for parents on managing conversations with children about the changes, equip children with strategies to stay connected safely through alternative platforms and prepare families for age assurance. Educational resources could include fact sheets, digital toolkits, advertisements, and interactive sessions. These resources would be co-designed and co-delivered with young people, educational groups and ethnic communities to ensure that they are effective and received well by the communities who would be the most affected by the Phase One legislation.
156. [BUDGET SENSITIVE] The regulatory function and DIA policy would be provided with additional resources under Budget 26 to stand up a permanent policy team that would have responsibility for assisting the implementation of Phase One as well as potential future phases of online safety regulation, and to expand the Digital Safety Group to enable it to fulfil the responsibilities of the Phase One regulator.

There are some risks with implementation

157. We have identified the following risks with implementation of the preferred options:

- **Circumvention by children:** Some children would likely find ways around the age assurance methods to create and hold accounts with age-restricted platforms. Circumvention is inevitable; however, we still consider the proposal would overall reduce the risk of online harm to children. This risk would be mitigated by educating young people about the proposal and why it is being implemented, as well as keeping child-appropriate online spaces open (such as YouTube Kids).
- **Shift to alternative or fringe services:** Some children may find other platforms to engage with their peers that are less safe or scrutinised. Migration to other platforms is a concern that would be managed similarly to circumvention, by educating children and keeping child-appropriate online spaces open. This risk would also be mitigated by ensuring the definition of an age-restricted platform is futureproofed to capture emerging and newly popular platforms that young people may shift to.
- **Personal information taken for age assurance may be misused:** There is a risk that platforms use personal information from age assurance to train their algorithms or in targeted advertising. We would place additional privacy protections on the collection and use of personal information for platform's fulfilling their minimum age requirements, tied to significant penalties, to ensure that personal information would not be misused.
- **Market exit by platforms:** Some platforms may decide to stop operating in the New Zealand market if the value of staying is outweighed by cost of complying with the new requirements. We are addressing this risk by aligning our proposal, where appropriate, with Australia where many platforms are already complying with minimum age requirements, and by ensuring requirements are not unnecessarily burdensome for platforms.

How will the proposal be monitored, evaluated, and reviewed?

158. In addition to enforcement and education responsibilities, the regulator would be required to monitor the efficacy of the Phase One legislation. The regulator would do this by testing how effective the age assurance methods used by platforms are. It could do this in a number of ways,^{6(c)} [REDACTED] and by requiring a platform's age assurance technology or child safety reporting processes to undergo an audit. The audit could be completed in-house by the regulator, or it could contract an independent third party, provided the third party had the necessary skills and did not have a conflict of interest.
159. The regulator, as part of its statutory responsibilities, would produce reports on platforms' compliance with Phase One requirements. These reports could include information on:
- the different types of age assurance technologies being used by platforms
160. how effective age assurance technologies were at preventing under 16s from holding an account
- whether platforms were complying with data privacy requirements for age assurance
 - which platforms had completed child safety reporting and where to find versions of the reports or summaries
 - any significant enforcement action the regulator had to take against a non-compliant platform.
161. The public could also report non-compliant platforms to the regulator, and this would only be available on the age assurance requirements. This could include parents or teachers who discover that a child in their care is able to make an account on an age restricted platform or members of the public who find out they have been interacting with a child on an age restricted platform. The regulator could act on reports it receives (which could include referring to the relevant platform) and take enforcement action if necessary to uphold Phase One requirements. It will not be within the regulator's remit to take action for users on online harms generally, such as complaints made to platforms on content moderation issues or content seen online. Reports would only be actioned if they are limited to Phase One requirements, and with a focus on minimum age requirements.
162. Social media and AI companion's impact on children is a globally recognised issue, and as such, there are many academics and research groups that are considering these impacts as well as the efficacy of legislation designed to address them. Officials anticipate that the impacts of Phase One legislation and global impacts of minimum age restrictions on platforms would be proactively assessed by both domestic and international academics.
163. [BUDGET SENSITIVE] As well, DIA could commission post-implementation research into the Phase One legislation as there is specific funding requested for an evaluation.
164. Prior to their minimum age requirements taking effect, Australia's eSafety Commission appointed an academic advisory group of eleven experts, led by Stanford University's Social Media Lab, to evaluate the implementation and outcomes of their legislation.³⁵

³⁵ eSafety Commissioner [eSafety appoints Stanford University-led academic advisory group to assess the impacts of the Social Media Minimum Age obligation](#) September 2025.

Officials spoke with the advisory group member from Stanford University during targeted consultation, who told us that their evaluation methodology could be replicated to assess New Zealand's legislation and they are collecting New Zealand baseline data.

165. To ensure that the Phase One legislation is reviewed post-implementation, it would contain a section requiring the Government to commence a review of the legislation's operation and prepare a report to be presented to the House of Representatives. Review clauses are common in newer legislation that implement novel proposals. For example, the Russia Sanctions Act 2022 requires that the Minister must review the legislation, as soon as practicable, two years after commencement. Officials consider within two years after commencement is a suitable timeframe for a review of the Phase One legislation. It is enough time for the policies to have embedded to be assessed, minimum age legislation from Australia to provide supporting evidence given its earlier implementation as well as consider the changes in technology that would have likely occurred since developing the Phase One policies. This will enable Phase Two options to be developed and considered following implementation of Phase One, as per Cabinet's agreement in December 2025.

Proactively Released

Appendix A – List of groups included in targeted consultation

Online safety regulators, overseas policy departments and independent Crown entities

1. Advertising Standards Authority
2. The Australian Department of Infrastructure, Transport, Regional Development, Communications, Sport and the Arts
3. Broadcasting Standards Authority
4. Classification Office
5. Coimisiún na Meán (Ireland online safety regulator)
6. eSafety Commissioner (Australia online safety regulator)
7. Netsafe
8. NZ On Air
9. Ofcom (United Kingdom online safety regulator)
10. UK Department of Science, Innovation and Technology

Māori groups

11. Representative from Māori Advisory Group to the Digital Identity Services Trust Framework Board
12. Representative from National Iwi Chairs Forum
13. Te Hāpori Matihiko

Youth and population organisations, including children and young people

14. Disabled Persons Assembly
15. End Child Prostitution, Child Pornography, and Trafficking of Children for Sexual Purposes (ECPAT NZ)
16. Gen Z Aotearoa (Youth-led national advocacy network for rangatahi)
17. InsideOUT
18. Le Va
19. Let Me Speak
20. Makes Sense
21. Save the Children Changemakers (group of 15 young people aged 15-20)
22. Triple P Parenting
23. Under-16s in rural communities (group of 10 children aged 12-17 via NZ Rural Schools Leadership Association)
24. VOYCE Whakarongo Mai
25. Young Farmers
26. Youth Parliament Digital Safety Working Group (former members aged 17-20)

Major online platforms and digital services, including representative bodies

27. Amazon
28. Apple
29. Discord
30. Google (also YouTube)
31. Kea Kids News
32. Kick
33. Meta
34. Microsoft
35. Music with Michal
36. Reddit

- 37. Snapchat
- 38. TikTok
- 39. Twitch
- 40. WeCreate
- 41. X

Gaming and Artificial Intelligence sector representatives

- 42. AI Forum
- 43. AI Safety Forum
- 44. Interactive Games and Entertainment Association
- 45. Roblox

Age assurance and digital infrastructure stakeholders

- 46. 2Degrees
- 47. Age Check Certification Scheme
- 48. Age Verification Providers Association
- 49. Digital Identity New Zealand
- 50. InternetNZ
- 51. My Mahi
- 52. N4L
- 53. NZ Tech
- 54. One NZ
- 55. Sush Labs
- 56. Spark
- 57. The Telecommunications Forum

Advocacy, rights, privacy and independent oversight bodies

- 58. B416
- 59. Chief Victims Advisor
- 60. Children's Convention Monitoring Group (consisting of representatives from: Te Kāhui Tika Tangata Rights Commission, Children's Rights Alliance New Zealand, Save the Children New Zealand, UNICEF Aotearoa)
- 61. Council for Civil Liberties
- 62. Human Rights Commission
- 63. Mana Mokopuna - Children's Commissioner
- 64. Office of the Privacy Commissioner

Academics and researchers, including thinktanks

- 65. Brainbox Institute
- 66. Electronic Privacy Information Centre
- 67. Koi Tū – Centre for Informed Futures
- 68. Reset Tech
- 69. Stanford Social Media Lab (leading eSafety's academic advisory group)

Health and wellbeing experts

- 70. Centre for Family Violence & Sexual Violence Prevention
- 71. Mental Health and Wellbeing Commission
- 72. Royal Australasian College of Physicians
- 73. Tu Te Akaaka Roa – Royal Australian & New Zealand College of Psychiatrists

Other

74. Anonymous individual 1

75. Anonymous individual 2

Proactively Released

Appendix B – International approaches to online safety

Country	Approach to online safety
United Kingdom	The United Kingdom operates a system-based duty of care framework under the Online Safety Act, requiring platforms to conduct children's access and risk assessments and put in place measures to mitigate identified harms. The Act includes strong age assurance obligations: platforms must use "highly effective age assurance" to prevent children from encountering harmful content such as pornography or self-harm material. In January 2026, the Government launched a three month public consultation on children's digital wellbeing, including potential minimum-age restrictions for social media, curfews, and limits on addictive design features.
European Union	The European Union uses a similar system based duty of care approach through the Digital Services Act (DSA), which requires platforms to assess and mitigate systemic risks to children, adapt service design to reduce harm, provide transparency around recommender systems, and implement proportionate age-appropriate protections. While the DSA does not set a minimum age for social media access, it establishes a regulatory baseline within which member states are now exploring additional measures. The EU Commissioner is now considering a minimum-age restrictions.
France (EU)	France's National Assembly approved key elements of a Bill introducing a minimum age of 15 for social media use in January 2026, with the Government aiming to enact legislation before the start of the September 2026 school year. ³⁶
Poland (EU)	In January 2026 the Polish Government announced plans for a minimum age of 16 for social platforms, supported by verification through its national digital identity wallet. Draft legislation is expected by February 2026. ³⁷
Spain (EU)	The Spanish Government has confirmed that it is working on legislation to ban social media use for under-16s and require age-verification systems. ³⁸
Denmark (EU)	Denmark is progressing legislation to ban social-media use for children under 15, with provisions allowing parental consent from age 13. The measure has cross-party support and could become law as early as mid-2026. ³⁹
Austria (EU)	The Austrian Government is considering a ban on social-media use for children under 14 and is currently exploring options for age verification. ⁴⁰
Ireland (EU)	The Irish Government has confirmed it will introduce legislation to ban social media access for under-16s, stating that while the preference is to move to restrict access at a European level, Ireland will take action regardless. ⁴¹

³⁶ Reuters, [France's National Assembly approves banning under-15s from social media](#), January 2026.

³⁷ Biometric Update, [Poland to use its digital ID app for age verification on social platforms | Biometric Update](#), January 2026.

³⁸ BBC, [Spain announces plans to ban social media for under-16s](#), February 2026.

³⁹ AP News, [Denmark plans to ban access to social media for children under 15](#), November 2026.

⁴⁰ Cybernews, [Austria and Poland plan social media bans for kids](#), January 2026.

⁴¹ The Irish Times, [New digital and AI strategy to confirm plans for under-16s social media ban](#), February 2026.

Norway	The Prime Minister of Norway has announced plans to introduce legislation raising the minimum age for social media use to 15, supported by secure age verification- mechanisms. ⁴²
Singapore	Singapore has introduced a layered online-safety regime combining app-store age-assurance requirements, platform-level safety obligations, and strengthened takedown powers. The Government is assessing whether to extend age-assurance requirements to social-media services, building on existing Codes of Practice for online safety. ⁴³
United States of America	In the United States, states are progressing a patchwork of age verification and parental consent laws for minors on social media- platforms, although implementation varies due to legal challenges. ⁴⁴
Canada	Canada is considering a national ban on social media access for children under 14 as part of upcoming online harms legislation and is developing a national age assurance- standard. ⁴⁵

For live updates on the implementation of social media age restrictions globally, see this tracker: [Global Social Media Age Restriction Tracker | Tech Policy Press](#).

⁴² The Guardian, [Norway to increase minimum age limit on social media to 15 to protect children](#), October 2024.

⁴³ Channel News Asia, [Govt to discuss age-appropriate content with platforms; social media cannot be a 'free-for-all'](#), August 2025.

⁴⁴ Age Verification Providers Association, [US state age assurance laws for social media](#), October 2025.

⁴⁵ NOW Toronto, [Canada considers social media ban for kids under 14](#), January 2026.

Appendix C – Additional evidence

Research	Key points
InternetNZ, ' New Zealand's Internet Insights ', 2024.	InternetNZ surveys show strong public concern about the impact of online harms on children, with 73 percent of adults identifying young children accessing inappropriate content as their top internet-related concern. Evidence of public concern
Analysis of data provided by CensusAtSchool on Oct 9, 2025. Includes responses from students aged under 16 from the 2021-22, 2023-24 and 2025 CensusAtSchool.	Students under the age of 16 were asked which platforms they had used in the past week. YouTube led with 78.1%, followed by TikTok 53.3%, then Snapchat on 52.4% and Instagram with 45.3%. Evidence of regular use of TikTok, Snapchat, YouTube and Instagram among under-16s.
Public Health Communication Centre Aotearoa, ' Evidence supports the need to protect adolescents from social media harms ' PHCC, July 2025. This is written by an academic who is a member of B416, an advocacy group.	Evidence in systematic reviews investigating the impact of high levels of use of social media on children suggests it has been associated with depression, anxiety, psychological distress, sleep disruption, eating disorders, and low self-esteem. Correlational evidence of social media use causing harm
Amy Orben et al., ' Windows of developmental sensitivity to social media ', Nature Communications, 2022.	There is substantial research to support that social media use is predictive of a subsequent decrease in life satisfaction for those under 16, who are developmentally more vulnerable. Evidence of developmental vulnerability for under 16s
Police Association, NZPA: ' Sextortion: It's not just blackmail ' 2024.	A 2024 press release from the Police Association reported on the prevalence of sextortion in New Zealand, particularly that the New Zealand Police deal with an average of 10 cases a week flagged as sextortion. Of the cases reported over a six-month period in 2024, 88 percent of the victims were male, and 18 percent were under the age of 17. Evidence of prevalence of sextortion in New Zealand
Education Gazette - ' Is social media weakening students' sense of belonging at school? ', 2019.	28% of New Zealand students in the Programme for International Student Assessment who were classed as 'extreme users' of the internet (averaging more than six hours a day on the internet outside of school hours) had much lower academic performance than students with lower internet usage. Correlational evidence of the negative impact of children's excessive online engagement on educational outcomes.

Youthline, [State of the Generation](#), 2023.

Research includes 13–15-year-olds and 16-24-year-olds.

Half of young people surveyed viewed social media as a fundamental issue facing their generation, particularly its effect on mental health and potential to cause social problems including increased levels of bullying. The research noted that while children are concerned about social media and its ability to spread misinformation and negatively impact relationships, body image and confidence, it is also the best way to reach them.

Evidence of youth-reported risks relating to mental health, bullying, and misinformation.

Internet Matters, [Me, myself & AI: Understanding and safeguarding children's use of AI chatbots](#), 2025

Vulnerable children reported even greater reliance on AI chatbots, with 71% of children identified as vulnerable using them regularly, 26% preferring them over real people, and 23% turning to chatbots because they lack someone else to talk to.

Evidence of heightened reliance on AI chatbots among vulnerable children.

Proactively Released