Unforeseen Consequences

How Child Safety Policy Risks Expanding Identity Exposure

A critique of the proposed *Children's Online Privacy Code*

The Problem We Think We're Solving

Over the past few years, the Australian government has taken an increasingly interventionist approach to the question of children's safety online. Public and academic concern about the influence of social media, exposure to harmful content, and the long-term effects of digital environments on young people has driven a wave of proposed regulation. At face value, this concern is both understandable and justified. Few would argue against the concept that children deserve meaningful protection in online spaces that are often opaque, algorithmically driven, and commercially incentivised in ways that do not align with their wellbeing.

It is within this context that policies such as social media age-verification requirements have emerged (Australian Government, 2024), alongside more expansive proposals like the *Privacy (Children's Online Privacy) Code 2026 (Exposure Draft),* developed by the Office of the Australian Information Commissioner (OAIC, 2026a). The Code applies to services "likely to be accessed by children", a definition that, while seemingly reasonable, raises immediate questions about the breadth of its application in practice.

These measures are framed as necessary safeguards – tools designed to prevent underage access, hold platforms accountable, and create a safer digital environment for young Australians. Politically, they are difficult to oppose without appearing indifferent to child safety – a reality that can discourage critical scrutiny. Practically, they demand much closer scrutiny. There is a critical distinction that must be made at the outset: a policy can be well-intentioned and still be fundamentally flawed. More importantly, it can fail not only in its effectiveness, but in ways that actively introduce new risks – risks that extend far beyond the population it seeks to protect.

From within the school environment, the limitations of these policies are not theoretical. Students continue to access platforms nominally restricted to them with ease, employing simple and widely understood workarounds. Age-gating mechanisms, far from acting as meaningful barriers, function more as superficial checkpoints – easily bypassed and inconsistently enforced. This gap between legislative intent (Australian Government, 2025) and lived reality (Dervisevic, 2025; Randall & Guest, 2026) is not a minor implementation issue; it is a structural failure.

At the same time, the proposed expansion of these approaches through the Children's Online Privacy Code signals a significant escalation. Language that applies to services "likely to be accessed by children" raises an immediate and unavoidable question: what part of the modern internet does not meet that threshold? In practice, such framing risks extending age-verification requirements far beyond social media platforms, into the broader digital ecosystem as a whole. The result is not simply

targeted child protection, but the potential normalisation of identity verification as a prerequisite for participation in online life.

This shift carries consequences that have not been adequately addressed. Systems designed to verify age at scale inevitably require the collection, processing, and storage of sensitive personal information. Expanding these systems across a wide range of services increases both the volume of data collected and the number of points at which it can be compromised. In an environment where even well-resourced organisations experience breaches, the assumption that such a system can be implemented securely is, at best, optimistic.

This essay argues that the current trajectory of online age-verification in Australia represents a convergence of three critical failures: it does not effectively prevent underage access; it expands regulatory scope in ways that are disproportionate and poorly defined; and it introduces significant privacy and cybersecurity risks for the broader population. In doing so, it risks undermining not only the safety of young people, but the integrity of the digital environment for Australians as a whole.

If the goal is to genuinely protect children online, then policy must be grounded in how digital systems actually function, how users actually behave, and how risk is actually distributed. Without that grounding, even the most well-intentioned measures risk becoming not safeguards, but liabilities.

Intent vs Reality – The Social Media Ban

Before turning to the proposed Children's Online Privacy Code, it is necessary to examine the policy model that has already been tested in public: the social media age-verification ban. This section considers the ban on its own terms. It begins with the stated intention behind the policy, then compares that intention with the practical reality of implementation, circumvention, expert warnings, and displacement. The point is not that the government's concern for children is misplaced. The point is that the chosen mechanism has already shown signs of failure, and those failures matter because the same logic now appears to be expanding into broader privacy regulation.

Well-Intentioned Theory and Practical Failure

The intent behind age-verification policies is, on its surface, straightforward: to prevent underage users from accessing platforms deemed inappropriate, and to reduce exposure to harmful content online. In policy terms, this reflects a broader effort to shift responsibility onto platforms, requiring them to enforce age-appropriate access and better safeguard younger users, framed as an intervention to protect young people from mental health risks associated with social media (Kaye, 2025).

In practice, however, the effectiveness of these measures is deeply questionable. Age-gating systems are not new; nor are the methods used to bypass them. Entering a false

date of birth, accessing shared or older accounts, and using basic tools such as Virtual Private Networks (VPNs) are all widely known and easily accessible workarounds (Croft, 2025; Gazy, 2026). Within school environments, these behaviours are not exceptional: they are routine; they are shared openly; they are normalised. Students continue to access restricted platforms with little difficulty, often with minimal technical knowledge required. Reporting on the policy's rollout, Reuters noted that "the country's social media feeds were flooded with comments from people claiming to be under 16" (Kaye, 2025), underscoring how quickly the restrictions were bypassed in practice.

This disconnect between design and outcome is not simply a matter of imperfect implementation. It reflects a more fundamental issue: these systems rely on forms of verification that are inherently difficult to enforce in open digital environments (Bellovin, 2025). Where enforcement is weak, compliance becomes optional. Where compliance is optional, restrictions lose their effectiveness.

More importantly, these limitations were not unknown. They were not obscure, speculative, or newly discovered after implementation. Nearly two decades ago, Gilbert (2008) warned that age verification was "not technically feasible" without infringing on the privacy of both children and adults, because adults would also need to be screened to prove they were not minors. The article also noted that a "savvy 12-year-old child" could misrepresent their age and obtain access to online services, before asking how reliable identification could be expected to succeed in cyberspace when comparable checks remain difficult even in the physical world. These are not minor implementation concerns; they are foundational problems. When contemporary experts continue to identify the same issues – circumvention, unreliable enforcement, privacy trade-offs, and difficulty implementing age verification at scale – the question is no longer whether the risks were foreseeable. The question is why policy continues to proceed as if they were not (Gilbert, 2008; Eltaher et al., 2025; Lueks et al., 2026; eSafety Commissioner, 2026).

Known Risks, Repeated Warnings

The problems now emerging from age-verification policy are not simply the result of poor rollout, platform reluctance, or adolescent defiance. Those factors matter, but they do not fully explain the failure. The deeper issue is that age verification has long been recognised as a policy tool with serious structural limits. Its weaknesses are not peripheral; they sit at the centre of the model itself.

At the technical level, the central promise of age verification is that platforms can reliably distinguish between children and adults online, and use that distinction to control access. In theory, this appears straightforward. In practice, it depends on assumptions that become fragile almost immediately: that users will provide accurate information; that documents, accounts, devices, or biometric signals can be reliably

tied to the person actually using the service; that platforms can enforce these systems consistently; and that young people will not adapt faster than the systems designed to restrict them. Each of these assumptions has already been challenged by the literature (Lueks et al., 2026).

Gilbert's 2008 analysis is valuable because it shows that these problems are not new. Rather than treating age verification as a simple technical fix, Gilbert described "many unresolved issues surrounding the concept of age verification", including privacy, data security, authentication errors, identity theft, and liability (Gilbert, 2008). Most importantly for the present debate, she identified the broader privacy problem created by the logic of age verification itself: "In order to protect children, it might be necessary to obtain information from everyone, so that a classification can be made" (Gilbert, 2008). That is the core issue. A system cannot reliably identify children without first assessing everyone.

More recent technical analysis reaches similar conclusions. Lueks et al. (2026) argue that "essentially all" age assurance technologies are vulnerable to technological circumvention, and that these circumvention methods place an upper limit on the effectiveness of any system. Age assurance, in this view, may increase friction, but cannot guarantee exclusion. The paper identifies multiple routes around these systems, including VPN use, adult-assisted access, and the transfer or sharing of verified accounts and devices. This matters because each workaround attacks the basic enforcement logic of age verification: the system may verify *someone, somewhere, at some point*, but it cannot reliably guarantee that the verified person is the child or adult actually using the service later.

This is where the problem shifts from technical limitation to policy failure. If a system is known to be vulnerable to circumvention, then enforcement cannot be treated as a simple matter of platform compliance. Platforms can be required to introduce checks. Regulators can threaten penalties. Governments can announce obligations. But none of this resolves the underlying problem that the system being enforced remains unreliable. At best, enforcement produces friction. At worst, it produces the appearance of protection while leaving the underlying access problem intact.

The privacy warnings are equally serious. Age verification systems do not operate in a vacuum; they require infrastructure. Depending on the method used, this may involve identity documents, biometric images, behavioural data, third-party verification services, device-level attestations, or persistent account-based tracking. Each method carries its own risks, but the shared problem is that verification creates pressure to collect, process, or infer more information about users than would otherwise be necessary (Alajaji, 2025; Lueks et al., 2026; OAIC, 2026a).

Lueks et al. (2026) describe these trade-offs clearly. Document-based age verification can create risks of exclusion, tracking, profiling, and information leakage to the verifier, because official identity documents often contain far more information than is required to establish age. The same paper warns that some age assurance architectures allow third parties to learn which services users are accessing, while other models create incentives for more persistent tracking, so users do not have to verify their age repeatedly. In other words, the privacy risk is not an accidental side effect. It is built into the attempt to make age verification usable at scale.

The policy danger, then, is not merely that age verification may fail to keep children away from restricted platforms. It is that the attempt to make it work may normalise more intrusive forms of identification and tracking across the internet. A weak system can still impose strong burdens. A bypassable checkpoint can still require everyone else to hand over sensitive information. A policy can fail as child protection while succeeding as infrastructure for surveillance, profiling, or identity collection.

These concerns should have demanded caution. Instead, age verification has continued to be presented as though the main challenge was implementation rather than design. That framing matters. If the problem is implementation, then the solution is more compliance, better technology, or stronger enforcement. But if the problem is design, then tightening the system may deepen the harm: more intrusive checks, more identity collection, more account-linking, more pressure on platforms to monitor behaviour, and more incentives for users to evade the system entirely.

The warnings were already there. They were present in older legal and technical analysis. They appear again in contemporary assessments of age assurance. They are visible in early evidence of circumvention and uneven compliance. The question, then, is not whether policymakers could have anticipated these risks. The question is why a policy model with known weaknesses continues to be expanded before those weaknesses have been resolved.

The Displacement Problem

A further problem with age-based access restrictions is that they do not operate evenly across the internet. They are most enforceable against large, visible, commercially exposed platforms with Australian legal presence, reputational risk, and regulatory obligations. These are the platforms most likely to respond to government pressure, build compliance systems, remove accounts, and publicly demonstrate cooperation. They are also the platforms most visible to parents, teachers, regulators, journalists, and researchers. In other words, they are the platforms most likely to be imperfectly controlled, but at least *partially* accountable.

The same cannot be said for the wider internet. The online spaces most capable of ignoring, avoiding, or resisting Australian regulatory expectations are often the very

spaces where oversight is weakest. Offshore services, anonymous forums, private servers, encrypted groups, fringe communities, and deliberately non-compliant platforms do not necessarily respond to Australian policy in the same way as mainstream commercial platforms. Some may lack an Australian presence. Some may be difficult to identify or regulate. Some may have little reputational incentive to comply. Others may actively frame non-compliance as part of their appeal (eSafety Commissioner, 2022; Cubitt & Morgan, 2024; Taylor, 2025).

This creates a displacement problem. If young people are pushed away from mainstream platforms, it does not necessarily follow that they will disengage from online social life. Some may simply create new accounts, use older accounts, borrow access, or use VPNs. Others may migrate towards platforms that are less visible, less moderated, and less accountable. The policy may therefore reduce access to regulated platforms while leaving less regulated spaces available, and in some cases more attractive, to the young people it is intended to protect. This concern has been echoed by UNICEF, which warned that "many children and young people will still access social media, whether through workarounds, shared devices, or turning to less regulated platforms, ultimately making it harder to protect them" (UNICEF, 2025).

UNICEF's warning underlines the fact that this risk is not hypothetical. This concern should be taken seriously. Mainstream social platforms are far from harmless, and there is a strong argument that their design choices have contributed to serious problems for young people. But they are not the only sources of online harm, and they are not necessarily the worst ones. A teenager displaced from Instagram or TikTok is not automatically displaced into safety. They may instead be displaced into online spaces with weaker moderation, fewer reporting tools, more anonymity, less adult visibility, and stronger exposure to harmful subcultures.

This is where sites such as 4chan become relevant. 4chan is not simply "another social media platform" in the ordinary sense. It is an anonymous imageboard with a long history of controversial, extremist, pornographic, racist, misogynistic, conspiratorial, and deliberately provocative content (Bernstein et al., 2011; Thorleifsson, 2021; Ling, 2023). It is also the kind of space that does not fit neatly into the public imagination of "social media harm", because it is not built around personal profiles, follower counts, influencer culture, or family-visible accounts. Yet for precisely that reason, it illustrates the weakness of a policy model focused on major platforms. If regulation removes young people from more visible social environments while leaving darker or more radicalising spaces untouched, the policy may not reduce harm at all. It may merely change where that harm occurs.

The same logic applies beyond any single website. The internet contains forums, message boards, gaming-adjacent communities, private Discord servers, encrypted chat groups, livestreaming spaces, file-sharing communities, and fringe platforms that

can host harmful content or facilitate harmful social dynamics. Some of these spaces are benign. Some are valuable. Some are essential for marginalised young people seeking community or support. But others are toxic environments where extremist ideology, conspiracy narratives, and content depicting or advocating violence do not merely circulate; they are reinforced, normalised, and sometimes actively promoted (Cubitt & Morgan, 2024).

A policy that concentrates enforcement on mainstream platforms therefore risks creating a misleading sense of success. Account removals, compliance reports, and public statements from major companies may give the appearance of action. They may even produce measurable reductions in underage accounts on specific platforms. But a reduction in visible underage accounts is not the same thing as a reduction in harm. If young people removed from those platforms remain online elsewhere, the policy has not solved the problem. It has moved the problem into spaces where it may be harder to see and harder to intervene.

This matters especially in schools. When students use mainstream platforms, the risks are not invisible. Teachers hear about conflicts that begin on Instagram. Parents notice Snapchat use. Peers talk about TikTok trends. Harm can still be missed, but it often leaves traces in social life that adults and institutions can eventually recognise. By contrast, movement into anonymous forums, private servers, or fringe online spaces may reduce that visibility. The harm does not disappear simply because it leaves the platforms most familiar to adults. It may become less legible, less reportable, and less reachable.

There is also a perverse incentive problem. The platforms most likely to comply are punished with compliance burdens, public scrutiny, and regulatory exposure. The platforms least likely to comply may become comparatively more attractive to users seeking fewer restrictions. This does not only matter for young Australians. When harmful communities become more attractive or more active, the risks can spill beyond the young users originally targeted by this policy. In this environment, responsible platforms become harder to use, while irresponsible spaces remain accessible. This is the exact opposite of what child protection policy should aim to achieve. A sound policy should raise safety standards across the ecosystem. It should not create incentives for users to migrate towards parts of the internet least interested in their safety.

None of this means that mainstream platforms should be exempt from responsibility. They should not be. Major platforms have profited from young users' attention, data, and social dependency, and they should be held accountable for design choices that amplify harm. But accountability must be distinguished from displacement. It is one thing to require platforms to improve safety, moderation, privacy settings, recommender algorithms, reporting pathways, and child-appropriate design. It is

another thing to assume that blunt exclusion from mainstream platforms will result in meaningful protection. Those are not the same policy.

The central flaw is that the ban treats access to particular platforms as though it were equivalent to exposure to harm. But harm is not confined to the platforms named in legislation or covered by enforcement attention. Harm moves. Users move. Communities move. Content moves. A regulatory approach that cannot account for this movement mistakes platform removal for child protection.

A policy that removes children from visible platforms but leaves them with access to darker, less moderated, and less accountable spaces does not eliminate risk; it redistributes it, and may, ultimately, exacerbate it. If young people are pushed from imperfectly regulated spaces into barely regulated ones, then the policy does not merely fail to protect them. It may make protecting them harder.

The Strongest Case for the Ban – and Why It Still Falls Short

The strongest argument in favour of the social media ban is not difficult to understand. Young people are experiencing real harms in online environments. Bullying, harassment, and social comparison are not imaginary concerns. Nor are parents, teachers, and policymakers wrong to worry about the power major platforms hold over young people's attention, identity formation, and social lives. Any serious critique of the ban must begin by acknowledging that the problem it claims to address is real.

Supporters of the ban may therefore argue that even imperfect protection is better than inaction. If age restrictions reduce the number of underage users on major platforms, even partially, that reduction may be seen as worthwhile. From this perspective, the policy does not need to be perfect to be justified. Seatbelt laws, alcohol restrictions, tobacco controls, and road-safety rules are not perfectly enforced either, yet they still shape behaviour, establish norms, and reduce harm.

This argument has force. The fact that a law can be broken does not, by itself, prove that the law is useless. Nor does the existence of workarounds automatically mean that regulation should be abandoned. A society can reasonably decide that certain activities are inappropriate for children and that platforms should not be allowed to profit from children's exposure to harm simply because enforcement is difficult.

However, the problem is not merely that the social media ban is imperfect. The problem is that its imperfections are not incidental. They are tied to the mechanism itself. If the policy fails only at the margins, then partial effectiveness may be enough. But if the policy's failures are structural – if circumvention is routine, enforcement is uneven, and compliance produces new privacy and displacement risks – then the case for "better than nothing" becomes much weaker.

Partial reduction is only a meaningful success if harm is actually reduced. A fall in visible underage accounts on major platforms may look like progress, but it does not necessarily show that young people are safer. If some young people leave mainstream platforms and disengage from harmful online environments, that is a genuine benefit. But if others move to less visible spaces, borrow access, use VPNs, or shift into unmoderated communities, then the policy may simply produce better-looking numbers while making harm harder to detect.

A second argument in favour of the ban is that platforms should be forced to take responsibility. This is also persuasive. Major platforms have designed systems that maximise engagement, collect user data, and monetise attention. They have often responded slowly or inadequately to harms affecting young users. It is reasonable to reject the idea that child safety should be left entirely to parents, schools, or individual children while platforms continue to profit from the environments they create.

But forcing platforms to take responsibility is not the same thing as forcing them to exclude children through age verification. Platform accountability could take other forms: stronger privacy defaults, limits on algorithmic amplification, restrictions on addictive design features, better reporting mechanisms, more transparent recommender systems, stronger moderation of harmful content, and meaningful consequences for platforms that knowingly expose children to foreseeable harm. These measures target the design of the environment itself. A blunt age ban, by contrast, targets access while leaving many of the underlying design incentives intact.

A third defence is that the ban may help shift social norms. Even if some young people bypass restrictions, the law may send a clear message: these platforms are not appropriate spaces for children. That message may empower parents, strengthen school policies, and give families a clearer basis for setting boundaries. In this sense, the law may be defended not only as an enforcement tool, but as a cultural signal.

Again, there is truth in this. Laws do not only punish behaviour; they also express societal values. But a symbolic law still needs to be judged by its consequences. If the cultural signal is achieved by building systems that normalise identity checks, increase pressure for data collection, and push some young people towards less accountable spaces, then the symbolism comes at a high cost. A law that sends the right message through the wrong mechanism can still produce harmful results.

Finally, supporters may argue that doing nothing is unacceptable. On this point, they are right. Doing nothing would leave young people exposed to platform designs that are often manipulative, commercially driven, and poorly aligned with their wellbeing. But “doing nothing” is not the only alternative to broad age-gating. The choice is not between total inaction and intrusive, unevenly enforced exclusion. A better policy response would focus on reducing harm where it is produced: in platform design, data

practices, recommender systems, advertising models, moderation failures, digital literacy gaps, and the lack of meaningful support for parents and schools.

The strongest case for the ban therefore identifies a real problem, but does not rescue the policy from its central weakness. Protecting children online is necessary. Holding platforms accountable is necessary. Establishing stronger norms around young people's digital wellbeing is necessary. But none of these goals require accepting a system that is easily bypassed by those determined to avoid it, unevenly enforced across the internet, and potentially dangerous for the privacy and security of everyone else.

The issue is not whether children deserve protection. They do. The issue is whether this mechanism protects them effectively enough to justify the risks it creates. On the evidence available, the answer is a clear and resounding no.

From Failed Gatekeeping to Identity Exposure

The social media ban matters not only because of its own weaknesses, but because it reveals the risks of building child-safety policy around age classification. Once that model is accepted, it becomes tempting to apply it elsewhere: to more services, more contexts, and more forms of online participation. The proposed Children's Online Privacy Code should be understood in that light. Although it is framed as a privacy measure rather than a ban, it may create new pressure for services to determine who is a child and who is not. This section examines how a policy designed to protect children's data may, in practice, encourage wider age assurance and expose users to greater identity risk.

The Failure That Should Have Prompted Caution

The argument so far is not that children face no real harm online, nor that governments should abandon attempts to regulate digital environments. The problem is more specific: the social media ban relies on a mechanism that is easy to bypass, difficult to enforce evenly, vulnerable to displacement effects, and burdened by privacy risks that have been known for years. It attempts to solve a real problem through a fragile gatekeeping method.

Given that evidence, the logical policy response should be caution. If age verification has not reliably accomplished what it set out to do, the next step should be to refine the approach, narrow its scope, examine its failures, and confront the privacy and cybersecurity risks it creates. Instead, the emerging direction appears to be broader. Rather than stepping back from a flawed model, the government now appears to be moving towards a framework that could extend age-assurance expectations across a much wider range of online services.

That is what makes the proposed *Privacy (Children's Online Privacy) Code 2026 (Exposure Draft*) so significant. On its face, the Code is not a social media ban. It is framed as a privacy measure: a way to ensure that online services recognise children's vulnerability, limit harmful data practices, and design digital environments with children's interests in mind. These goals are legitimate. The concern is not the stated purpose of the Code. The concern is the mechanism by which that purpose may be pursued, and the breadth of the services to which it may apply.

The key phrase is "likely to be accessed by children". At first glance, that sounds sensible. A law designed to protect children online should obviously apply to services children are likely to use. But once that phrase is applied to real people accessing the modern internet, its boundaries become difficult to locate. Children do not only access social media. They access games, search engines, streaming services, forums, shops, news websites, educational tools, messaging services, cloud platforms, app stores, video platforms, fan communities, productivity tools, and general websites. In practice, there are very few online services that can confidently say they are not likely to be accessed by a child.

This is the escalation. The social media ban attempted to build a gate around selected platforms. The Children's Online Privacy Code risks extending the logic of the gate to the internet more broadly. If services must identify whether a user is a child in order to determine what obligations apply, then age assurance becomes not an exception for high-risk platforms, but a potential compliance layer across ordinary online life. With wider use comes a wider attack surface. The more surfaces required to assess age, the more systems will collect, process, infer, or transmit age-related data. If billion-dollar companies cannot always protect basic login data, it is reasonable to ask whether every smaller service drawn into this compliance framework can be trusted with identity documents, biometric checks, or other sensitive age-assurance data.

The danger is not only that such a system may fail. The danger is that it may fail while becoming normal.

What the Code Actually Says

The proposed *Privacy (Children's Online Privacy) Code 2026* is not, on its face, another social media ban. It is a proposed privacy code made under the *Privacy Act 1988*, intended to regulate how covered entities handle children's personal information. The Explanatory Statement frames the Code as a response to concerns that existing privacy protections are inadequate for children online, noting that children are increasingly reliant on "online platforms, social media, mobile applications and internet-connected devices", and that these services "routinely collect and use large volumes of personal information about children (OAIC, 2026b).

That concern is legitimate. The explanatory statement also notes growing concern about children being “datafied”, including through the collection of information about their “activities, interests, location, wellbeing and relationships” (OAIC, 2026b). In that sense, the Code is aimed at a real and serious problem: the routine extraction, processing, profiling, and commercial use of children’s personal information in digital environments.

The stated purpose of the Code is to require covered entities to adopt “best practice approaches to the handling of children’s personal information”, increase transparency, recognise the value of children’s privacy, centre the “best interests of the child”, and provide children with greater control over their personal information (OAIC, 2026b). Those are not trivial goals. A framework that limits unnecessary data collection, restricts manipulative consent practices, improves children’s privacy notices, and gives children stronger rights over their own information is, in principle, a worthwhile reform.

The concern begins with scope. The Code does not apply to every online service in the most literal sense. Its application is tied to providers of a “social media service”, “relevant electronic service”, or “designated internet service”, using definitions drawn from the *Online Safety Act 2021* (OAIC, 2026a). However, within those categories, the Code applies to services that are “likely to be accessed by children” or “primarily concerned with the activities of children” (OAIC, 2026a). That distinction matters. The Code is not written as a universal identity law, but it may still reach far beyond the public’s ordinary understanding of “social media”.

The Explanatory Statement confirms this broader application. It states that the Code applies to APP entities providing social media services, relevant electronic services, or designated internet services where the service is “likely to be accessed by children”, and also to certain services “primarily concerned with the activities of children” (OAIC, 2026b). It gives examples of services primarily concerned with children’s activities that may not be directly accessed by children themselves, including applications that track early childhood development, family photo-sharing applications, online school management systems, and internet-connected baby monitors (OAIC, 2026b).

The most important operational provision is section 8. The Code states that, before a covered entity collects personal information about an end-user, it “must take steps that are reasonable in the circumstances to ascertain the age of the end-user" (OAIC, 2026a). This is the point at which a child privacy framework begins to intersect with age assurance. If an entity must know whether it is dealing with a child before it collects personal information, then the entity needs some way of assessing age.

The Explanatory Statement makes that link explicit. It explains that an entity may take reasonable steps to ascertain an end-user's age “by implementing age assurance”, and describes age assurance as an umbrella term covering “age inference, age estimation,

age verification, self-declaration, and parental attestation" (OAIC, 2026b). In other words, the Code does not mandate one single technical method, but it does create a legal setting in which covered services may need to choose some method of age assessment.

There is an important caveat. Section 8 does not apply if the entity applies the Code's child protections to the service "regardless of age of end-users" (OAIC, 2026a). In theory, that means an online service can avoid age assurance by treating all users as though they are entitled to the Code's protections. This is a meaningful safeguard, and it should be acknowledged. However, it does not eliminate the concern. For many services, applying child-specific rules to every user may be commercially, technically, or functionally unattractive. Where that is the case, age assurance becomes the practical alternative.

The Explanatory Statement also confirms that the intended approach is risk-based. It says lower-risk services may accept a higher degree of uncertainty about an end-user's age, while higher-risk services would require a higher degree of certainty (OAIC, 2026b). That sounds reasonable, but it also embeds age assurance into ordinary compliance decision-making. Services will be expected to consider the risks created by their data practices, the available age assurance methods, their efficacy, cost, and "data and privacy implications on end-users" (OAIC, 2026b).

This is why the Code deserves close scrutiny. Its stated purpose is privacy protection, not surveillance. Its language is careful, risk-based, and in many places genuinely protective of children. But the compliance logic is clear: if a service is covered by the Code and does not apply child protections to everyone, it may need to take reasonable steps to ascertain users' ages. The Code does not need to say "everyone must upload ID" for an age-assurance infrastructure to emerge. It only needs to create enough legal pressure for services to decide that checking, estimating, inferring, or otherwise classifying age is the safest compliance path.

The issue, then, is not whether the Code contains good privacy principles. It does. The issue is whether those principles are being attached to a compliance model that may extend age assessment across a much wider portion of online life. A law designed to limit children's data exposure could, if implemented carelessly, produce a new incentive for services to collect or infer more information about everyone in order to decide who receives those protections.

When "Likely" Becomes Almost Everything

The phrase "likely to be accessed by children" is doing a great deal of work. At first glance, it appears modest. It suggests a reasonable limit: the Code applies not to the entire internet, but to online services children are likely to use. In isolation, that sounds appropriate. A children's privacy code should not be confined only to services explicitly

designated for children, because many children use services designed for general audiences. The problem is that, once applied to the modern internet, the phrase becomes difficult to contain.

The Code itself does not use the phrase in isolation. It applies to the provision of a "social media service", "relevant electronic service" or "designated internet service" that is "likely to be accessed by children" or "primarily concerned with the activities of children" (OAIC, 2026a). The Explanatory Statement also emphasises that the Code applies to specific services, rather than necessarily every service offered by an entity. It gives the example of a bank whose pocket-money app may be captured while its business banking or home-loan app may not be (OAIC, 2026b). This qualification matters. It prevents the argument from being overstated. The Code is not written as a blanket rule for every activity of every organisation.

But that does not remove the practical problem. The example of the bank's home-loan app is useful precisely because it shows what a relatively clear exclusion looks like. Some services are genuinely unlikely to be accessed by children: business banking portals, professional tax software, wholesale supplier platforms, enterprise procurement systems, industrial compliance tools, corporate payroll systems, or specialised professional services aimed at adults in regulated industries. These are services with narrow functions, adult user bases, institutional access controls, or obvious practical barriers for child use. In such cases, a provider may be able to say with some confidence that the service is not likely to be accessed by children.

The difficulty is that much of the public internet does not look like that. Children do not occupy a separate, neatly fenced-off version of online life. They use search engines, video platforms, games, messaging tools, forums, shopping sites, streaming services, education platforms, app stores, cloud services, music platforms, fan communities, productivity tools, coding resources, news sites, image-hosting services, and general websites. Many of these services are not designed primarily for children. Many are not marketed to children. But that is not the test. A service may be general-audience, adult-oriented, or simply indifferent to age, and still be accessed by children.

This is where "likely" becomes unstable. It is not the same as "targeted at children", "designed for children", or "primarily used by children". It asks whether children are likely to access the service at all. That is a broader and more uncertain question. For a small business, forum operator, app developer, game server host, video platform, marketplace, or online community, the compliance question may not be "Are we a children's service?", it may become "Can we safely say children are unlikely to access this?". Those are very different questions.

This distinction matters because uncertainty changes behaviour. A service provider that knows it is outside scope can ignore the Code. A service provider that knows it is inside

scope can design for compliance. But a service provider that is unsure may behave conservatively. If the consequences of being wrong include regulatory scrutiny, reputational damage, legal advice, redesign costs, or complaints, many providers may decide that the safest course is to assume children might access the service and act accordingly. In practice "likely to be accessed by children" may become less a clear boundary than a compliance warning attached to much of the public-facing internet.

This is especially significant because the Code's obligations are not merely symbolic. Section 8 requires covered entities, before collecting personal information about an end-user, to take steps that are reasonable in the circumstances to ascertain the age of that end-user, unless the entity applies the Code's protections regardless of age (OAIC, 2026a). That means scope uncertainty does not merely affect legal classification. It may affect the design of ordinary access, account creation, privacy settings, consent flows, and data collection systems.

For large platforms this may be expensive but manageable. They can hire privacy lawyers, compliance teams, user-experience designers, policy staff, trust-and-safety teams, and age-assurance vendors. For smaller services, the pressure may be different. They may lack the resources to make subtle legal and technical distinctions about likely child access. Faced with uncertainty, they may adopt blunt compliance tools: self-declared dates of birth, age gates, reduced functionality, third-party verification systems, or broad restrictions on access. Some may treat all users as children. Some may decide to block younger users entirely. Some may simply collect more information than they previously needed to demonstrate that they took "reasonable steps".

This creates a paradox. A privacy code designed to reduce unnecessary collection of children's personal information may, under conditions of uncertainty, create incentives for more services to ask age-related questions of more users. The Explanatory Statement does recognise that entities can avoid age-assurance steps by applying the Code's protections to all end-users, and that lower-risk services may accept greater uncertainty. Those safeguards matter. But they do not erase the practical incentive. If a service wants to distinguish between children and adults, or if it believes its risk profile demands more certainty, it must find a way to classify users by age.

This is why the phrase "likely to be accessed by children" deserves more scrutiny than it has received. The public may understand the Code as applying to children's services, social media platforms, or apps obviously directed at young people. But the language reaches further than that. It captures services likely to be accessed by children, not merely services created for them. In a digital environment where children are ordinary users of the general internet, that difference is significant.

The question, then, is not whether the Code literally captures every online service. It does not. The question is whether its wording creates a broad grey zone in which many public-facing services may reasonably fear they are covered. That grey zone is where overcompliance happens. It is where legal uncertainty becomes design practice. It is where "likely" begins to operate, in practice, like "possible enough that we had better check."

This is the path from child protection to identity exposure. The danger is not that every website will immediately demand a passport. That would be an exaggeration. The danger is subtler and more likely: more services asking more users to declare, estimate, verify, or otherwise prove their age because the safest compliance position is to classify first and ask questions later. In that environment, "likely to be accessed by children" may not formally mean "the whole internet", but it may still push ordinary online life in that direction.

The Age-Assurance Ladder

The Code does not prescribe one single method for determining a user's age. This is important. It does not say that every covered service must demand a driver's licence, passport, birth certificate, facial scan, or government digital ID. Instead, the Explanatory Statement describes "age assurance" as an umbrella term covering a range of methods, including "age inference, age estimation, age verification, self-declaration, and parental attestation" (OAIC, 2026b). It also states that the approach is intended to be risk-based: lower-risk services may accept a higher degree of uncertainty, while higher-risk services may require a higher degree of certainty. In determining what steps are reasonable, entities are expected to consider available methods, their relative efficacy, implementation costs, and the "data and privacy implications on end-users" (OAIC, 2026b).

At first glance, this flexibility appears reassuring. A risk-based system seems preferable to a blunt universal identity check. But the available methods do not sit on equal ground. They sit on a ladder. At the bottom are methods that are simple, cheap, and relatively privacy-light, but easy to bypass. At the top are methods that may provide stronger assurance, but require more intrusive forms of identification, biometric processing, behavioural profiling, or third-party verification. The central trade-off is therefore difficult to avoid: the less intrusive the method is, the easier it is to defeat; the harder it is to defeat, the more intrusive it tends to become.

At the lowest end of the ladder is self-declaration. This is the familiar model: a user enters a date of birth, ticks a box, or states that they are above a particular age. It is cheap, easy to implement, and does not necessarily require the user to submit identity documents or biometric information. But as a meaningful safeguard, it is weak. Eltaher et al. describe self-declaration as the most common age-assurance method used across major online platforms, but note that it has "significant limitations" because

minors can falsify their age and because it does not involve additional assurance steps. Lueks et al. go further, treating self-declaration as a case where “no age assurance is enforced” because it does not technically identify minors (Eltaher et al., 2025; Lueks et al., 2026).

This matters because self-declaration can satisfy the appearance of a gate without providing the function of one. It creates a screen, not a safeguard. A child who wants access can simply enter a different birth year. Eltaher et al.’s testing of major social media platforms found that account-creation systems often relied on self-reported age, and their broader analysis classifies self-declaration as the “Asserted” level of age assurance: the least stringent form, with no checks to verify the accuracy of the asserted age and no formal authentication of the user’s identity (Eltaher et al., 2025).

A slightly stronger approach is age inference. This does not necessarily ask the user to prove their age directly. Instead, a platform or service attempts to infer age from account-based or behavioural signals. These may include browsing history, social media interactions, uploaded content, search queries, gameplay patterns, device metadata, or other traces of online activity. Lueks et al. define age inference as a method that deduces a user’s age from “digital behaviour, metadata, or other signals”, and note that some platforms already use AI-driven inference to flag accounts that may belong to underage users (Lueks et al., 2026).

This may appear less invasive than asking for a passport, but that appearance is misleading. Age inference depends on profiling. It makes the user’s behaviour the evidence. Bellovin’s discussion of Google’s behavioural age-estimation model is useful here: Google describes its model as using account-associated signals such as search history and categories of YouTube videos watched to determine whether a user is likely over or under 18. Bellovin notes that such signals can produce boundary problems because behaviours associated with age are not age itself; for example, searches about university admission may be made by older non-traditional students, younger students who skipped grades, or people in different national systems (Bellovin, 2025).

The deeper concern is not only accuracy, but incentive. Lueks et al. warn that age inference may normalise surveillance, increase data collection, push users towards account creation, and require sufficient accumulated data before it can work effectively. If age inference becomes a preferred compliance method, services may have a stronger reason to collect, retain, and analyse behavioural data. A method that avoids asking for identity documents may therefore still reduce privacy by making ongoing observation more normal (Lueks et al., 2026).

Age estimation sits further up the ladder. This may include tools that estimate age from a face scan, voice sample, body movement, or other biometric or quasi-biometric signal. Such systems may appear attractive because they seem to avoid formal identity

documents: the user does not need to prove exactly who they are, only that they are likely above or below a threshold. Lueks et al. describe facial age estimation as the most common form, where a user's face image is analysed by a machine-learning system to estimate whether they appear to be a minor or an adult (Lueks et al., 2026).

But this does not remove the privacy concern. A face scan is still sensitive. It is still captured, transmitted, analysed, and judged by a system the user may not understand. Nor is it necessarily reliable. Lueks et al. rate age estimation as "low-medium" assurance and note that facial age estimation performs worst around teenagers and young adults, precisely the group most relevant to age checks. They cite NIST testing showing mean absolute errors of around three to five years for teen subjects, meaning a 17-year-old might be classified as over 20 or as 13 to 14 in some cases. They also identify risks of bias and argue that biometric data should not generally be treated as privacy-friendly merely because it avoids formal identity documents (Lueks et al., 2026).

Parental attestation is another rung on the ladder. In some contexts, it may make sense for a parent or guardian to confirm a child's age or provide consent for the handling of personal information. The Explanatory Statement itself contemplates parental involvement, stating that if a child is under 15, consent must be obtained from a person with parental responsibility, and that entities must take reasonable steps to confirm that the person giving consent actually has that responsibility. It gives examples such as email communication, vouching through a bank, school or telecommunications provider, video conference confirmation, or government digital ID (OAIC, 2026b).

That may be workable in some situations, but parental attestation is not a universal solution. It assumes that a safe, available, digitally capable adult exists and is willing to participate. It also assumes that involving that adult is always in the child's interests. Lueks et al. note that parental control and consent can be effective only if parents actively and correctly set up the relevant systems. They also identify risks for children whose parents do not care, cannot manage the systems, or where the child has special needs or interests that they may not safely disclose to a parent, including LGBTQ+ young people seeking support (Lueks et al., 2026).

At the top of the ladder sit more serious identification systems: driver's licences, passports, birth certificates, government digital identity systems, bank verification, telco verification, or other trusted third-party identity checks. These may be more reliable than self-declaration or casual estimation. Lueks et al. describe age verification as relying on official sources of information, either through direct document-based verification or through an existing relationship with an identity provider such as a bank, government agency, health insurer, or mobile service provider (Lueks et al., 2026).

These methods may provide stronger assurance, but they are also far more consequential. Once a service asks for this level of proof, the issue is no longer just whether a user is old enough. The issue becomes what sensitive identity material is being collected, who processes it, whether it is stored, whether it can be linked to the user's activity, which third parties are involved, and what happens when something goes wrong. Lueks et al. warn that document-based verification can leak information to the verifier because the verifying party may learn the contents of an official identity document, and often receive a photo or video of the person. That information can go far beyond what is necessary to establish age (Lueks et al., 2026).

Gilbert made a similar point nearly two decades earlier. Age verification, she argued, is likely to require more than age alone: it requires identification and authentication. That means collecting attributes such as name, address, date of birth, and other identifying information, often with the involvement of identity providers or other third parties. Gilbert also warned that, to protect children, it may be necessary to obtain information from everyone in order to classify users as adults or minors. The result is a privacy problem that affects the whole user population, not only children (Gilbert, 2008).

This is the core trade-off. The methods least likely to expose sensitive identity information are often the easiest to defeat. The methods most likely to establish age with confidence are the methods most likely to expose users to privacy, security, and identity risks. A date-of-birth field can be bypassed by a child in seconds. A driver's licence check may be harder to bypass, but it asks the user to place far more trust in the service, its vendors, its databases, its security practices, and its future behaviour.

That trade-off matters because the Code's logic does not exist in a vacuum. It sits in a broader policy environment already experimenting with age gates and platform restrictions. If weak methods are adopted, they may satisfy paperwork while doing little to protect children. If strong methods are adopted, they may create precisely the identity exposure this essay is concerned about. Either way, the public should not be reassured merely by the phrase "age assurance". The term is broad enough to include both a meaningless date-of-birth box and a serious identity-verification process.

This is why the ladder matters. It prevents the debate from collapsing into false simplicity. Age assurance is not one thing. It is a spectrum of methods with different costs, risks, weaknesses, and consequences. The more the law pressures services to know the age of users, the more services will have to choose where on that ladder they sit. Some will choose light-touch methods that children can easily evade. Others will choose stronger systems that expose users to deeper identity risks. Both outcomes are troubling for different reasons.

The danger is not that every service will immediately choose the most intrusive option. The danger is that legal pressure, risk aversion, and compliance uncertainty will

gradually normalise age classification as a condition of access. Once that happens, the question becomes not only whether a particular method is accurate, but whether Australians should be expected to repeatedly declare, estimate, prove, infer, or verify their age simply to participate in ordinary online life.

The less intrusive the system is, the easier it is to bypass. The harder it is to bypass, the more intrusive it becomes. That is the unresolved tension at the centre of age assurance, and it is why a privacy framework built around age classification may end up creating the very identity exposure it should be trying to avoid.

When Verification Data Becomes Breach Data

The concern with identity-based age assurance is not simply that it is inconvenient. It is that the more seriously a system attempts to verify age, the more sensitive the data involved tends to become. A date-of-birth field is weak because it can be falsified. A face scan, driver's licence, passport, Medicare number, birth certificate, digital identity credential, bank check, or third-party identity service may be stronger, but only because it brings the user closer to formal identification. At that point, age assurance stops being a light-touch privacy measure and becomes part of the identity infrastructure of everyday online life.

This is the central cybersecurity problem. A weak age gate may not work. A strong age gate may create a valuable target. If services are encouraged, pressured, or required to determine whether users are children, then more services may collect, process, transmit, infer, or verify age-related information. Some may only ask users to self-declare. Others may rely on behavioural inference or facial estimation. Others may involve identity documents or third-party verification providers. Each additional system creates another place where sensitive data can be mishandled, over-retained, linked, breached, sold, subpoenaed, or repurposed.

Technical assessments of age assurance repeatedly identify this trade-off. The Knight-Georgetown Institute's technical assessment found that the most common age-assurance systems require users either to identify themselves by name, email, or phone number, or to provide an age verification provider with an image of their face. In doing so, the user is forced to trust that provider not to misuse the data, not to disclose it, and not to lose it in a breach (Rescorla, Arnao & Cooper, 2026). The report also concluded that systems with stronger technical privacy guarantees, such as device-based enforcement or zero-knowledge proofs, are possible but are not yet widely deployed (Rescorla, Arnao & Cooper, 2026).

The problem is not only what the service provider learns, but what the wider verification ecosystem learns. In a common web-based age-assurance model, the user is redirected from the service they want to access to an age verification provider. The user then provides age signals, such as a government ID, selfie, or live video, before being

redirected back to the original service. The KGI report's diagram of a typical web-based age assurance system illustrates this clearly: the age verification provider sits between the user and the service, evaluates the user's age signals, and then tells the service whether access should be granted (Rescorla, Arnao & Cooper, 2026). This means that the age verification provider may learn who the user is, what service they are trying to access, and when the check occurred.

That architecture matters because privacy does not depend only on promises. It depends on design. If a system technically prevents unnecessary information from being collected, linked, or retained, the user is safer. If the system instead collects sensitive information and relies on contractual assurances, privacy policies, deletion rules, and vendor trust, then the user must simply hope that every organisation in the chain behaves properly and remains secure. The KGI report makes this distinction directly: technical controls are generally stronger than policy controls, because users can often verify that technical controls exist, while policy controls require trust in organisations the user may have no relationship with and no meaningful ability to audit (Rescorla, Arnao & Cooper, 2026).

The Age Check Certification Scheme's report for the UK Information Commissioner's Office makes the same practical point from a measurement and compliance perspective. It notes that age-assurance providers require personal data to provide their services, and that "the greater the need for accuracy, the greater the quantity or sensitivity of personal data is needed" (Allen et al., 2022). It also warns that holding more personal data increases the importance of appropriate security measures, particularly where more sensitive categories of data are involved (Allen et al., 2022). That is the ladder again: higher confidence tends to pull more sensitive data into the system.

The problem becomes especially clear with document-based verification. A driver's licence, passport, Medicare card, or birth certificate is not simply an age token. It contains additional identity material. The Age Check Certification Scheme report uses the example of a UK driving licence, noting that while the document contains date-of-birth information, it also contains other personal information not necessarily relevant to age assurance, such as home address and vehicle licence categories (Allen et al., 2022). The same logic applies in Australia. A document that proves age often proves much more than age.

The *Trustworthy Age Assurance* report reaches a similar conclusion. It argues that document-based age verification may establish age with a higher degree of certainty, but does so by relying on hard identifiers and trusted records that can identify the user. The report warns that widespread age assurance raises risks including privacy intrusion, data leakage, behavioural surveillance, identity theft, exclusion, and discrimination (Sas & Mühlberg, 2024). It also concludes that no currently available

method adequately protects fundamental rights across the board, particularly when assessed against privacy, security, inclusivity, and circumvention risks (Sas & Mühlberg, 2024).

Australia does not need to imagine what happens when identity-rich datasets are compromised. It has already happened. In the Optus breach, the Queensland Government reported that the leak affected between 2.5 million and 9.7 million records, and that for some customers the exposed information included addresses, driver licence numbers, Medicare numbers, or passport numbers (Queensland Government, 2022). The point is not that Optus was an age-verification system. It was not. The point is that large organisations already hold identity material at scale, and when that material is compromised, the harm reaches far beyond an ordinary username-and-password leak.

The Latitude Financial breach makes the identity-document problem even clearer. Reporting in The Guardian stated that 14 million customer records were stolen, including 7.9 million Australian and New Zealand driver's licence numbers and 53,000 passport numbers (Barrett, 2023). That is precisely the kind of data stronger verification systems may ask users to provide: licences, passports, and other documents that can be used to establish identity. Once exposed, these details can feed scams, fraud, identity theft, and long-running administrative burdens for affected individuals.

Medibank provides a further warning because it shows that sensitive exposure is not limited to identity documents alone. ABC News reported that the personal information of 9.7 million Medibank customers was stolen in the 2022 breach, including names, dates of birth, Medicare numbers and sensitive private health information, and that much of it was published on the dark web (Crowley, 2024). Again, the point is not that age-verification systems would collect health records. The point is that well-resourced Australian organisations have already failed to protect highly sensitive personal information once it was collected and stored.

It is important to note that these breaches are not isolated outliers. They are the most visible examples of a much broader pattern. In the first half of 2025 alone, the OAIC received 532 notifiable data breach reports, and described breaches as remaining at a high level despite a decrease from the previous reporting period (OAIC, 2025). That figure captures only notifiable breaches reported under the scheme; it does not include every security incident, attempted breach, unreported compromise, or smaller privacy failure. The point is not that age-assurance systems will necessarily reproduce Optus, Latitude, or Medibank. The point is that Australia already has a large and continuing data-breach problem, even before policy creates new incentives for more identity-related information to circulate through online services.

These examples matter because they undermine the comforting assumption that sensitive verification data will simply be handled safely. Organisations routinely promise that personal information will be protected. They promise that documents will be deleted. They promise that third-party vendors are secure. They promise that access will be controlled. Sometimes those promises are sincere. But sincerity is not security. Once sensitive identity material enters a system, it depends on that system's architecture, retention rules, access controls, vendors, staff training, logging practices, incident response, and future business decisions. It also depends on whether attackers find one overlooked weakness.

The OAIC's own public guidance reflects the seriousness of this kind of exposure. Its advice for people affected by a data breach distinguishes between different categories of compromised information, including contact information, financial information, health information, tax-related information, and government-issued identity documents. Where a breach involves government-issued identity documents such as a driver licence or Medicare details, the OAIC advises affected people to contact the agency that issued the document for advice (OAIC, n.d.). This is useful not because it proves that any particular age-assurance system will fail, but because it shows that government-issued identity exposure is treated as a serious practical risk requiring follow-up action.

Nor is the harm of identity exposure comparable to a normal password leak. A password can be changed. A driver's licence number may sometimes be replaced, but not without cost, administrative burden and delay. A passport can be cancelled and reissued, but that does not undo the exposure. A date of birth cannot be changed. A face cannot be changed. A history of identity-document compromise can follow a person through credit checks, account openings, scams, phishing attempts and future fraud. This is why identity data is different. Once exposed, it cannot be cleanly recalled.

This risk is magnified if age assurance becomes normal across a wide range of services. A single carefully designed, privacy-preserving system may be defensible in a narrow, high-risk context. But a broad compliance environment can produce a patchwork of systems: large platforms building their own tools; smaller services buying off-the-shelf verification products; third-party providers handling checks across multiple services; and less sophisticated operators implementing whatever appears cheapest and easiest. The risk is not only one national database. It is a distributed identity-risk ecosystem, where sensitive information is copied, transmitted, transformed, logged, and retained across many organisations.

This is especially concerning for smaller services. Large companies at least have the resources to employ privacy lawyers, security engineers, compliance teams and external auditors, and they still suffer serious breaches. Smaller organisations may have fewer resources, weaker governance, and greater dependence on third-party

vendors. The Age Check Certification Scheme report makes this point in practical terms, noting that it would be unreasonable for a start-up providing basic age-estimation services to have the same security measures as a global organisation processing millions of age verifications (Allen et al., 2022). That may be reasonable from a proportionality perspective, but it also confirms the deeper problem: the security environment will be uneven.

The defender of identity-based age assurance may respond that good systems do not need to store identity documents. In principle, that is true. Privacy-preserving age assurance is possible. Cryptographic credentials, selective disclosure, device-based enforcement, double-blind transmission methods, and zero-knowledge proofs may reduce the amount of information revealed to services. Both the KGI and *Trustworthy Age Assurance* reports recognise that more privacy-preserving approaches are possible, and that these may allow users to prove age eligibility without disclosing full identity (Rescorla, Arnao & Cooper, 2026; Sas & Mühlberg, 2024).

But the existence of better designs does not guarantee their deployment. Privacy-preserving systems are technically complex, require careful implementation, depend on trustworthy issuers and software environments, and are not yet the ordinary reality of age assurance. KGI states that stronger privacy designs such as device-based enforcement and zero-knowledge proofs are possible, but not widely deployed (Rescorla, Arnao & Cooper, 2026). *Trustworthy Age Assurance* similarly concludes that promising privacy-preserving techniques are still developing and face implementation challenges, including security, inclusivity, feasibility, and the absence of a comprehensive legal and technical framework (Sas & Mühlberg, 2024). “Privacy-preserving” is not a magic phrase. It is a design standard, and design standards can be weakened, avoided, misunderstood, or implemented badly.

This is where the policy problem becomes clearest. The Children’s Online Privacy Code is framed around protecting children’s privacy, and many of its principles are legitimate. But if the practical response is to encourage more services to classify users by age, then the Code may create new incentives for identity exposure. Weak systems may fail to protect children. Strong systems may create valuable stores or flows of identity data. Poorly implemented systems may do both at once.

The question, then, is not whether any age-assurance system can ever be designed safely. Some can be designed more safely than others. The question is whether a broad, risk-averse compliance environment will reliably produce the safest possible systems across the full range of services likely to be captured. Australia’s recent data-breach history gives little reason for confidence. If some of the country’s largest and best-resourced organisations have failed to protect sensitive personal information, it is difficult to accept, without much stronger safeguards, that a wider ecosystem of online services should be trusted with more of it.

The danger is not only that verification data may be breached. The danger is that, in the name of protecting children, Australia may normalise the routine production of breach-worthy identity data in the first place.

The Best Defence of the Code – And Why It Is Not Enough

The strongest defence of the Children's Online Privacy Code is that it is not, in itself, a universal identity law. It does not simply order every Australian to upload a driver's licence before using the internet. It does not prescribe one single age-assurance method. It does not require all services to choose the most invasive option. In fact, the Code and its Explanatory Statement contain several safeguards that deserve to be taken seriously.

The first defence is that the Code is risk-based. Covered entities are not asked to treat every service, every user, and every data practice as equally dangerous. The level of certainty required when ascertaining age is meant to depend on the circumstances, including the risk of harm associated with the collection, use, or disclosure of personal information (OAIC, 2026b). In principle, this is sensible. A low-risk service should not be required to use the same age-assurance method as a high-risk service handling sensitive information or exposing children to significant harm.

But a risk-based framework does not remove the incentive problem. It simply moves the pressure into risk assessment. Services still need to decide whether they are covered, whether children are likely to access them, whether their data practices create risk, whether child-specific obligations apply, and what steps are "reasonable in the circumstances". For large platforms, that may mean a lengthy internal compliance process. For smaller services, it may mean uncertainty, legal advice they cannot afford, or the adoption of crude off-the-shelf tools. A risk-based system can be more proportionate than a blunt rule, but it can still produce overcompliance when the cost of being wrong is high.

The second defence is that the Code allows services to avoid age assurance by applying child protections to all users. This is an important caveat. If a service simply gives every user the Code's protections, it may not need to distinguish between children and adults in the first place (OAIC, 2026a). In privacy terms, that is often the better direction: minimise collection, reduce profiling, improve transparency, and design safer defaults for everyone.

This should be taken seriously as a possible path forward. If the Code pushes services to reduce unnecessary data collection for all users, it could produce genuine public benefit. Many adults would also benefit from clearer privacy notices, less manipulative design, tighter limits on profiling, and stronger controls over personal information. In that form, the Code's logic could move privacy law in the right direction: not by forcing

users to prove who they are, but by forcing services to justify why they collect so much information in the first place.

The problem is that this path may not be the one many services choose. Applying child-level protections to all users may be commercially unattractive for platforms built around behavioural advertising, personalised recommendation, engagement maximisation, or broad data collection (OAIC, 2026b). If adult users are more profitable when they can be tracked, profiled, targeted, nudged, and analysed more aggressively, some services will have a strong incentive to distinguish adults from children rather than raise the privacy floor for everyone. In that context, age assurance becomes the mechanism that preserves the existing adult data economy while carving children out as a protected category.

The third defence is that age assurance does not necessarily mean document verification. This is also true. Age assurance can include self-declaration, parental attestation, age estimation, age inference, device-level signals, third-party checks, digital credentials, or privacy-preserving proofs (OAIC, 2026b; Rescorla, Arnao & Cooper 2026). It would be misleading to pretend that the only possible outcome is a passport upload on every website.

But the range of methods is exactly the problem. The weaker methods are easier to bypass. The stronger methods are more intrusive. If a service uses self-declaration, it may do little more than recreate the same ineffective gatekeeping already seen in social media restrictions. If it uses behavioural inference, it may push services towards profiling. If it uses facial estimation, it may introduce biometric processing, accuracy problems, bias, spoofing, and false classification. If it uses document-based verification, it creates identity exposure. The fact that services have options does not eliminate the trade-off; it distributes the trade-off across the internet (Sas & Mühlberg, 2024; Rescorla, Arnao & Cooper, 2026).

The fourth defence is that privacy-preserving age assurance may solve much of this. In theory, this is the most compelling response. Properly designed systems could allow a user to prove that they are over or under a threshold without revealing their name, date of birth, address, document number, or browsing activity. Device-based enforcement, selective disclosure, double-blind systems, and zero-knowledge proofs may reduce the amount of personal information exposed during age assurance (Sas & Mühlberg, 2024; Rescorla, Arnao & Cooper, 2026). If such systems were mature, widely deployed, independently audited, accessible, interoperable, and legally constrained, they could answer some of the concerns raised here.

But that is a large “if”. The existence of better technical designs does not guarantee that they will be adopted, adopted well, or adopted evenly (Sas & Mühlberg, 2024; Rescorla, Arnao & Cooper, 2026). A privacy-preserving system can be undermined by poor

implementation, weak governance, vendor concentration, inadequate audit, unclear retention rules, bad user experience, exclusion of people without suitable devices or credentials, or commercial pressure to collect more data than necessary. The question is not whether a safe version can be imagined. The question is whether a broad compliance environment will reliably produce that safe version across the full range of services likely to be captured.

The fifth defence is that doing nothing is not acceptable. This is true, and it should be said plainly. Children do face real risks online. They can be exposed to harmful content, exploitative design, manipulative recommendation systems, unwanted adult contact, harassment, commercial profiling, and data practices they cannot reasonably understand. A serious critique of age assurance should not become an argument for regulatory inaction.

But rejecting a flawed mechanism is not the same as rejecting child protection. The question is not whether children should be protected. They should be. The question is whether this model of protection actually works, whether it is proportionate, whether it creates new risks, and whether those risks fall only on the platforms being regulated or on the public as a whole. A policy can be motivated by child safety and still expose children and adults to new harms.

The sixth defence is that imperfect protection is better than no protection. This argument has political force because it sounds practical. No law is perfect. No safeguard stops every bad actor. Seatbelts do not prevent every road death, but they still save lives. On this view, age assurance does not need to be flawless; it only needs to reduce harm enough to justify its costs.

The problem is that this comparison only works if the costs and benefits point in the same direction. If age assurance merely blocks some underage access while displacing other children to less regulated spaces, normalising broader identity checks, increasing profiling, and creating new stores or flows of sensitive personal information, then the policy may not simply be imperfect. It may be counterproductive. Partial effectiveness is not enough if the measure also creates systemic risks that are larger, longer-lasting, or harder to repair.

A final defence is that regulation can force the market to improve. Under this view, age-assurance technology may be imperfect now, but legal requirements will push providers towards better systems, stronger standards, certification, privacy-preserving methods, and more responsible implementation (Allen et al., 2022; Sas & Mühlberg, 2024; Rescorla, Arnao & Cooper, 2026). There is some truth in this. Regulation can create demand for safer tools. It can raise expectations. It can force companies to take children’s privacy seriously.

But regulation can also entrench bad systems if it moves faster than the technology, the standards, and the oversight needed to make those systems safe. A mandate does not automatically create trustworthy infrastructure. It can create a market for compliance products before the public has had a serious debate about the risks (Sas & Mühlberg, 2024; Rescorla, Arnao & Cooper, 2026). It can reward vendors that offer fast, cheap, legally convenient tools rather than those that best protect privacy and security. It can also shift responsibility away from platform design and onto users, who are then asked to prove their age repeatedly while the underlying business models remain largely intact.

These counterarguments matter because they show that the issue is not simple. The Code contains legitimate aims. It includes safeguards. It allows flexibility. It may encourage better privacy practices in some contexts. It is not, by itself, a crude command for universal identity checks. Any fair critique must recognise that.

But those concessions do not answer the central concern. A broad, risk-based, flexible framework can still create broad, risk-averse, identity-exposing behaviour in practice. If services believe they need to know who is a child in order to comply, many will seek ways to classify users by age. If weak methods are accepted, the protection may be illusory. If strong methods are used, the privacy and cybersecurity risks increase. If privacy-preserving methods are promised but not reliably implemented, users are left trusting systems they cannot see, audit, or meaningfully refuse.

The best defence of the Code is that it seeks to protect children without mandating one universal method of identification. The weakness in that defence is that legal pressure does not need to mandate universal identification directly in order to normalise it indirectly. It only needs to make age classification feel like the safest path.

That is why the Code should not proceed on reassurance alone. It needs clear limits, stronger safeguards, independent technical scrutiny, strict data-minimisation requirements, meaningful enforcement against overcollection, protection for anonymous and pseudonymous access, and a genuine preference for applying privacy protections to all users where possible. Without those constraints, a privacy code designed to protect children may become another step towards identity exposure as a condition of ordinary online life.

The System This Creates

The preceding sections have dealt with two related problems: first, that age-verification policy is unreliable as a child-safety mechanism; and second, that expanding age-assurance expectations may create new privacy and identity risks. Taken together, these problems point towards a broader question: what kind of system does this produce? The answer is not a clean, child-safe internet. It is a fragmented system in which determined users continue to find workarounds, compliant services absorb

heavier burdens, non-compliant spaces remain available, and ordinary users become increasingly accustomed to proving or declaring themselves before participating online.

Workarounds Do Not Disappear

The first thing this system creates is not certainty, but adaptation. Users do not encounter age checks as immovable barriers. They encounter them as obstacles, and obstacles invite workarounds. This is especially true for young people, who are often highly motivated, socially connected, and quick to share practical methods for getting around restrictions. A policy that assumes age assurance will simply separate children from adults misunderstands how users behave in real digital environments.

The available workarounds are not exotic. Some are barely technical at all. A child may enter a false date of birth, use an existing account, borrow an older sibling's login, ask a parent to verify access, share a device, or use an account created before restrictions were introduced. Others may use VPNs, overseas services, alternative app stores, private links, mirror sites, or technical tools shared through peer networks. These methods vary in sophistication, but they share the same basic effect: they weaken the relationship between the person verified by the system and the person actually using the service.

This matters because age assurance often verifies a moment, not an ongoing reality. A system may confirm that someone, somewhere, at some point, appeared to meet an age threshold. It may not reliably confirm that the same person continues to use the account, device, browser, or session. If an adult verifies a device and then hands it to a child, the system may still see a compliant adult user. If a child uses an older account, the platform may see an account history that predates the restriction. If a young person uses a VPN to appear outside the relevant jurisdiction, the service may never apply the Australian rule at all. The legal category may be clear, but the technical enforcement is messy.

The result is a familiar pattern: the people most determined to evade the system are often the least burdened by it, while ordinary users, cautious users, and compliant services absorb the friction. A teenager who wants to bypass a restriction may try three methods in ten minutes. An adult who simply wants to access a forum, game, video platform, or online service may be asked to declare, estimate, or verify their age before continuing. The system therefore risks becoming most effective against the cooperative and least effective against the determined.

This does not mean that all restrictions are meaningless. Friction can change behaviour. Some young people may be stopped. Some may be delayed. Some may decide the effort is not worth it. For younger children especially, even a weak barrier may reduce casual access. That limited effect should be acknowledged. But friction is not the same as protection, and a policy that relies on friction should be honest about

what it can and cannot achieve. It may reduce some access at the margins; it does not create a secure boundary between children and harmful online spaces.

This is particularly important once the logic expands beyond a small number of platforms. If age assurance becomes more common across online life, workarounds will not remain isolated tricks. They will become shared habits. Young people will not simply learn how to bypass one platform. They may learn how to bypass the category of age checks itself: which VPNs work, which sites do not ask, which platforms are less strict, which verification methods are easiest to fool, which accounts can be borrowed, and which services can be accessed through alternative routes. The broader the system becomes, the more valuable the knowledge of evasion becomes.

There is also a risk that workarounds push users towards more dangerous behaviour. A child who lies about their date of birth has not necessarily exposed themselves to additional harm beyond the platform itself. But a child who searches for bypass methods, downloads unfamiliar tools, uses questionable VPNs, follows advice from anonymous forums, or borrows credentials may encounter new risks. A policy intended to keep children away from harm can inadvertently teach them to hide their activity more effectively, move further from adult visibility, and rely on tools or communities that adults are less able to supervise.

The central point is simple: workarounds do not disappear because Parliament wishes them away. They are part of the environment in which these laws operate. Any serious child-safety policy must begin from that reality. If a system is designed around the assumption that users will comply honestly, it will fail against the users most motivated to avoid it. If it is tightened to prevent evasion, it becomes more intrusive for everyone else. That tension has not been resolved. It has merely been pushed downstream, onto platforms, verification providers, families, schools, and users themselves.

The Compliant Are Punished; the Non-Compliant Are Rewarded

The second consequence is that the burden of the system falls unevenly. Laws of this kind are easiest to enforce against services that are visible, established, commercially exposed, and willing to operate within Australian law. Large platforms can be contacted, investigated, fined, reported on, and publicly pressured. They have legal teams, trust-and-safety departments, compliance staff, and reputations to protect. If the government tells them to build an age-assurance system, they are the services most likely to attempt it.

That may sound like the system working. In one sense, it is. Regulated platforms should not be allowed to ignore child safety simply because compliance is difficult or expensive. But the problem is that these are not the only spaces young people use, and they are not necessarily the spaces least interested in safety. A platform that attempts to comply may become more frustrating, more restrictive, and more identity-heavy for

ordinary users, while services that ignore the rules may remain comparatively easy to access.

This creates a perverse incentive. The services that try to obey the law take on the highest compliance burden. They may be forced to redesign account creation, adjust data flows, integrate third-party age-assurance providers, limit features, introduce verification steps, or collect new information to show that they have taken “reasonable steps”. These changes cost money, create friction, and introduce privacy risks. They may also make the service less attractive to users who do not want to be classified, verified, or interrupted.

By contrast, a non-compliant service may offer the path of least resistance. It may not ask for identity documents. It may not block access. It may not remove underage users. It may not care about Australian regulators, Australian parents, Australian teachers, or Australian law. It may operate offshore, avoid clear ownership, resist moderation, or treat defiance as part of its culture. From the user’s perspective, especially a young user trying to bypass restrictions, that can make the non-compliant space more appealing, not less.

The result is not a cleaner internet. It is an uneven one. Compliant services become heavier, more cautious, and more bureaucratic. Non-compliant services remain lighter, faster, and easier to enter. If the most responsible actors are made harder to use while the least responsible actors remain accessible, then the policy may reward the very behaviour it should discourage. It punishes the services willing to build safeguards and leaves untouched the services least likely to provide them.

This matters for privacy as well as safety. A compliant platform may implement age assurance because it is legally exposed. That system may require the user to declare an age, submit a face scan, connect an account, use a digital credential, or interact with a third-party verifier. A non-compliant platform may require none of this. The privacy-conscious adult, the determined teenager, or the user who simply dislikes identity checks may therefore be nudged away from regulated services and towards less accountable ones. The system can make the safer path feel more invasive and the riskier path feel easier.

There is also a market effect. Compliance costs do not fall equally. Large platforms may absorb them. Smaller services may struggle. A major company can negotiate with vendors, build internal compliance systems, and survive regulatory scrutiny. A small forum, indie game community, niche educational tool, hobby platform, or start-up may not be able to do the same. Some may over comply. Some may block users. Some may shut down features. Some may collect more information than they otherwise would because the simplest available compliance product tells them to. Others may decide that Australian users are not worth the risk.

This creates a strange hierarchy of burden. Large compliant platforms can survive, but become more invasive. Small compliant services may be pressured into crude or excessive systems. Non-compliant services may simply continue. The user, meanwhile, is left navigating a fragmented environment where the amount of friction they face depends less on the level of harm and more on whether a particular service is visible enough, resourced enough, or legally exposed enough to comply.

That is not a stable foundation for child protection. A sound policy should make harmful behaviour harder and responsible behaviour easier. It should raise the baseline of safety without making lawful, ordinary participation more fragile. But if the effect of the system is to impose the greatest burdens on those willing to follow the rules, while leaving evasive or hostile spaces relatively untouched, then the policy is not merely incomplete. It is structurally misaligned.

None of this means compliant platforms deserve sympathy for having to protect children. They do not. Many large platforms have profited enormously from user attention, data extraction, and addictive design. They should face real obligations. But those obligations should target the harms they create: recommender systems, profiling, manipulative design, weak moderation, poor reporting pathways, and unsafe defaults. A compliance model built primarily around age classification risks shifting attention away from those design failures and onto the user's obligation to prove eligibility.

The central problem is that the system can mistake visible compliance for meaningful safety. A platform may build an age gate and still remain harmful. A service may collect less identity data and still be safer. A fringe site may ignore age assurance entirely and continue exposing users to harmful content. The presence of a checkpoint is not the same as the presence of protection.

In practice, this system may produce an internet where the easiest services to regulate become the most burdened, and the hardest services to regulate become the least changed. That is a poor bargain. It asks compliant services and ordinary users to absorb the cost of a system that may leave the most dangerous, least accountable spaces largely intact.

Displacement Into Worse Spaces

The third consequence is displacement. If young people are blocked, removed, or discouraged from using mainstream platforms, it does not follow that they simply go offline. Some will. Some may spend less time online, and that may be a genuine benefit for those individuals. But others will move sideways. They will find platforms that ask fewer questions, use accounts that are harder to trace, join private spaces, or shift into online communities where adult oversight, platform moderation, and regulatory pressure are weaker.

This is one of the central weaknesses of gatekeeping as a child-safety strategy. It treats access to particular platforms as though those platforms are the problem itself. But the online harms affecting young people are not confined to the major platforms named in public debate. Harmful content, grooming risks, extremist material, misogynistic subcultures, self-harm content, conspiratorial thinking, scams, pornography, harassment, and exploitation do not disappear because a teenager loses access to Instagram, TikTok, Snapchat, Facebook, YouTube, Discord, Steam, or X. Those risks exist across a wider ecosystem.

Mainstream platforms are deeply flawed. They can amplify harmful material, reward outrage, enable harassment, and monetise attention in ways that are damaging to young people. But they are also visible. Parents know their names. Teachers hear about them. Regulators can contact them. Journalists can investigate them. Researchers can study them. Users can report content. There are public expectations, however imperfectly met, that these platforms should moderate abuse, respond to complaints, and maintain safety systems.

Less regulated spaces often lack those same pressures. Anonymous forums, fringe communities, private servers, encrypted groups, mirror sites, offshore platforms, and deliberately non-compliant services may be harder for parents, teachers, regulators, and researchers to see. They may also be more hostile to moderation itself. Some spaces build their identity around being unfiltered, ungoverned, anti-mainstream, or resistant to outside scrutiny. For a young person pushed away from a visible mainstream platform, those spaces may not be safer alternatives. They may be more dangerous ones.

The problem is not merely that young people might encounter “bad content” somewhere else. The problem is that displacement can change the social environment around that content. On a mainstream platform, harmful material may appear in a feed, be challenged by peers, reported by users, moderated by staff, or at least exist within a wider mix of content. In more isolated or extreme spaces, the same material may be reinforced. It may become part of group belonging. It may be wrapped in irony, humour, grievance, identity, or rebellion. It may be repeated until it feels normal.

This is especially concerning for students who are socially isolated, angry, vulnerable, curious, or seeking community. A young person locked out of one online space may not interpret that as protection. They may interpret it as exclusion. If the alternative spaces they find provide belonging, status, attention, or an explanation for their frustration, the risk is not only exposure but recruitment into a worldview. Harm does not merely move; it can deepen.

This does not mean that every private server, forum, gaming community, or anonymous space is dangerous. Many are harmless. Some are genuinely supportive. Some provide

marginalised young people with friendship, advice, humour, shared interests, and a sense of identity they may not find offline. A serious policy critique should not treat the entire wider internet as a moral panic. The point is narrower: if policy pushes young people out of spaces that are visible and at least partially accountable, it must consider where they may go next.

The displacement problem therefore exposes a gap between political simplicity and social reality. It is politically simple to say that children should not be on certain platforms. It is much harder to build systems that reduce harm across the places young people actually move through. Young people do not experience the internet as a list of regulated categories. They experience it as a set of pathways: links, chats, recommendations, invitations, memes, servers, usernames, group chats, search results, and peer suggestions. If one path is blocked, another can appear quickly.

A child-safety policy that cannot account for those pathways may produce false confidence. It may reduce visible underage activity on some platforms while increasing hidden activity elsewhere. It may make parents and politicians feel that something has been done, while making the actual behaviour of young people harder to observe. In schools, this matters. Harm that happens in visible peer spaces is already difficult to manage. Harm that moves into anonymous, private, or fringe spaces becomes even harder to identify, interrupt, or discuss.

The aim should not be to preserve mainstream platforms as they currently exist. They need reform. The aim should be to avoid a policy design that mistakes platform exclusion for harm reduction. If the system makes young people less visible without making them safer, it has failed. If it moves them from flawed but accountable spaces into less accountable ones, it may do more than fail. It may make intervention harder.

The Normalisation of Proving Yourself

The fourth consequence is cultural as much as technical. A society can become accustomed to things that once would have seemed excessive. The first age check may feel unusual. The second may feel irritating. The tenth may feel normal. Over time, the expectation changes: access to online life becomes something that must be preceded by declaration, classification, estimation, verification, or proof.

That shift matters because the internet has long allowed different levels of identity. Sometimes people use their real names. Sometimes they use pseudonyms. Sometimes they browse anonymously. Sometimes they create accounts only for a specific purpose. Sometimes they participate quietly without being forced to attach their offline identity to every action. This flexibility has value. It allows people to explore ideas, seek help, participate in communities, read sensitive information, ask embarrassing questions, organise politically, and move through digital spaces without making every action part of a permanent identity trail.

Age-assurance systems do not necessarily abolish that flexibility all at once. The risk is more gradual. If more services decide that they need to classify users by age, then more users become accustomed to being sorted before participation. Some systems will ask for a date of birth. Others may ask for a face scan. Others may rely on account history, behavioural inference, device signals, government credentials, or third-party verifiers. The individual interaction may seem minor. The cumulative effect is not minor.

The deeper danger is that classification becomes the default. Before reading, posting, messaging, watching, joining, searching, or buying, the user may first need to be placed into a category. Adult or child. Verified or unverified. Eligible or ineligible. Low-risk or high-risk. Trusted or untrusted. Once that becomes normal, the burden of proof shifts onto the user. Participation is no longer presumed. It is granted after classification.

This is the point at which a child-safety measure begins to reshape the broader digital environment. The public debate may focus on children, but the infrastructure necessarily touches adults as well. Adults must prove they are adults. Teenagers near the threshold must prove they are old enough, or young enough, depending on the rule. Children may be asked to identify themselves as children in order to receive protection. Families may be asked to verify parental responsibility. Everyone becomes part of the sorting process.

This is not simply an inconvenience. It changes the relationship between the user and the internet. A person who must repeatedly prove eligibility may become more cautious about where they go, what they read, what they ask, and what communities they join. They may avoid lawful material because the verification step feels invasive. They may avoid sensitive services because they do not trust the verifier. They may decline to participate because the cost of access feels too high. The result is not only friction, but a quieter narrowing of ordinary online freedom.

It also changes expectations around privacy. If users are repeatedly told that identity or age checks are normal, they may become less likely to question why those checks are needed, what information is collected, who receives it, how long it is kept, and whether a less invasive alternative exists. This is how privacy erosion often works. Not through one dramatic collapse, but through repeated small demands that gradually reset what people are willing to tolerate.

The irony is that this is being done in the name of privacy. The Children’s Online Privacy Code is intended to limit the exploitation of children’s personal information. That aim is legitimate. But if the practical outcome is a broader expectation that users must be classified before they can participate, the policy risks building a privacy problem into the mechanism of privacy protection. It may reduce certain forms of data use for children while normalising new forms of data collection, inference, or verification for everyone.

There is a better path. Services can be required to collect less data. They can be required to design safer defaults. They can be restricted from profiling children. They can be required to justify recommender systems, dark patterns, manipulative notifications, and harmful engagement loops. They can be made more accountable for the environments they build. None of that necessarily requires making identity proof the default doorway into online life.

The normalisation of proving yourself should therefore be treated as a serious policy cost. It may not appear in the same way as a breach, a fine, or a compliance report. But it changes the baseline of digital citizenship. It teaches Australians that access comes after verification, that privacy is conditional, and that participation depends on being legible to the system.

That is a dangerous lesson. A safer internet should not require Australians to become more identifiable at every turn. It should require powerful services to become less exploitative, less opaque, and less careless with the people who use them.

What Better Policy Would Look Like

Criticising age assurance does not require accepting the internet as it currently exists. Children do face real harms online, and many of those harms are intensified by platform design, commercial incentives, algorithmic amplification, weak moderation, and poor privacy practices. The answer is not inaction. The answer is a policy framework that addresses the conditions producing harm, rather than placing the primary burden on users to prove their age or identity before participating online.

A better approach would begin from a different premise: children should be protected because online services are made safer, less exploitative, and less data-hungry, not because every user is forced through a classification system. Where age assurance is genuinely necessary, it should be narrow, proportionate, independently scrutinised, and designed to minimise identity exposure. But it should not become the default solution to every online safety problem.

Start With Platform Design, Not User Identification

The first priority should be platform design. Many of the harms associated with social media and online services are not caused simply by the presence of children on those platforms. They are caused by systems designed to maximise attention, encourage compulsive use, amplify emotionally charged content, recommend increasingly extreme material, and convert user behaviour into commercial data.

If that is where harm is produced, then that is where regulation should focus. Governments should be asking harder questions about recommender systems, algorithmic transparency, engagement-maximising design, manipulative notifications, dark patterns, addictive interface loops, inadequate reporting systems, and weak

moderation practices. Platforms should be required to demonstrate that their systems do not foreseeably expose children to harm, rather than being allowed to treat age gates as a substitute for safer design.

This distinction matters. A platform can verify age yet remain harmful. A service can exclude children and still operate an exploitative model for everyone else. Conversely, a service can be made safer by reducing profiling, limiting harmful recommendation pathways, improving reporting tools, and designing for user wellbeing without requiring every user to hand over identity information. Child safety should not be reduced to the question of who gets through the gate. It should include the design of the space on the other side.

Raise the Privacy Floor for Everyone

A better privacy framework should raise baseline protections for all users. The Code already gestures towards this possibility by allowing services to apply child protections regardless of age. That idea should be taken seriously and strengthened.

Rather than requiring services to identify children so that children alone can receive better privacy protections, policymakers should ask which of these protections would *not* benefit Australians more broadly. Limits on unnecessary data collection, stronger transparency requirements, restrictions on manipulative consent flows, tighter controls on profiling, and meaningful user rights over personal information would benefit children and adults alike.

This approach avoids one of the central problems identified in this essay. If services must distinguish children from adults to decide who receives protection, they may be pushed towards age classification. But if the privacy floor is raised for everyone, the need to classify users is reduced. In many contexts, the safer question is not “how do we identify the children?” but “why is this service collecting so much personal information from anyone?”.

This would also reflect the reality that adults are not immune to exploitation, manipulation, profiling, or breach risk. Children deserve protection, but many harmful data practices are not acceptable simply because the target is over 18. A policy that improves privacy for everyone is less likely to create identity exposure and more likely to address underlying business models that make online spaces unsafe.

Keep Age Assurance Narrow, Necessary, and Proportionate

Age assurance should not be treated as inherently illegitimate. There may be contexts where it is necessary: genuinely high-risk services, adult-only content, gambling, or situations where legal access restrictions already exist and are clearly justified. But its use should be narrow, necessary, and proportionate.

That means age assurance should not be used merely because it is administratively convenient. It should be used only where there is clear evidence of need, where less intrusive measures are insufficient, and where the expected safety benefit outweighs the seriousness of the risk. A low-risk service should not be pushed towards high-assurance identity checks simply to create regulatory comfort.

Where age assurance is used, strict safeguards should apply. Services should collect the minimum information necessary, avoid retaining identity documents, prefer privacy-preserving methods, provide meaningful alternatives for users who cannot use a particular method, and ensure that users can challenge errors. Third-party verification providers should be independently audited, subject to strong breach obligations, and prohibited from reusing verification data for advertising, profiling, analytics, or unrelated commercial purposes.

Most importantly, age assurance should not become a general-purpose compliance shortcut. It should remain exceptional. If the only way to make a policy work is to make users more identifiable across ordinary online life, then the policy needs to be narrowed.

Support Schools, Parents, and Young People Directly

A better approach must also support the people who actually live with these issues every day. Schools, parents, carers, and young people cannot be treated as afterthoughts in online safety policy. They are often the first to notice the consequences when digital harms spill into classrooms, friendships, families, and mental health.

Schools need practical resources, not just slogans. Digital literacy curriculum outcomes should include more than warning about screen time. Students need to understand algorithms, privacy, misinformation, scams, extremist recruitment, image-based abuse, persuasive design, data trails, and the social pressures that shape online behaviour. Teachers need usable curriculum support, clear reporting pathways, and realistic guidance that reflects how young people actually use technology.

Parents and carers also need support, but policy should not assume every household has the same time, knowledge, confidence, language background, digital access, or safety. Some parents will be highly engaged. Others may be overwhelmed, absent, controlling, unsafe, or digitally excluded. Systems that rely heavily on parental consent or parental enforcement can therefore reproduce existing inequalities. Support should include accessible education, practical tools, community-based programs, and clear advice that does not simply shift responsibility onto families.

Young people themselves should be consulted seriously. Too often, online safety policy speaks about children while treating them as passive objects of protection. But young people can understand the pathways, pressures, workarounds, and social

consequences of online life in ways adults often do not. If policy ignores their behaviour, it will continue to design systems for imaginary users.

Children cannot be protected from the internet only by locking doors. They also need to be taught how to recognise danger, ask for help, protect their privacy, understand manipulation, and move through digital spaces with increasing independence. Protection and education need to work together.

Safeguards Before Expansion

Finally, any expansion of age assurance should be paused until stronger safeguards are in place. The current trajectory asks Australians to trust that broad, risk-based, flexible compliance frameworks will not become broad, risk-averse identity-exposing systems in practice. That trust has not been earned.

Before expanding age-assurance expectations, policymakers should define the scope of covered services more clearly. The phrase “likely to be accessed by children” should not be allowed to operate as an open-ended compliance warning across ordinary online life. If the intention is not to capture the wider internet, then the law should say so clearly. If certain categories of services are meant to be excluded, those exclusions should be explicit.

There should also be independent technical scrutiny before age-assurance systems are deployed at scale. Systems should be assessed for effectiveness, privacy, security, accessibility, bias, circumvention resistance, and breach risk. These assessments should not rely solely on vendor claims or political reassurance. They should be public enough to support democratic accountability while protecting legitimate security details.

Stronger data-minimisation rules are essential. Services should not collect identity documents, biometric data, behavioural profiles, or persistent identifiers unless there is a clear and proportionate justification. Where sensitive data is processed, retention should be strictly limited, reuse should be prohibited, and breaches should trigger serious consequences. Users should not be forced to trust vague promises that data will be handled responsibly.

There should also be explicit protection for anonymous and pseudonymous access where identification is not necessary. The ability to read, search, learn, ask questions, join communities, and access lawful information without constantly providing identity is not a minor convenience. It is part of a healthy digital society.

The goal should be clear: protect children without building unnecessary identity infrastructure. That means resisting the temptation to treat age classification as the default answer to every online safety problem. It means recognising that privacy, security, access, and child protection are not separate concerns. A policy that protects

children by exposing everyone else to new identity risks has not solved the problem. It has relocated it.

A Safer Internet Cannot Be Built on Fragile Gates

The desire to protect children online is not misplaced. It is one of the few points in this debate that should not be controversial. Children do face real risks in digital environments. They can be exposed to harmful content, exploitative design, manipulative recommendation systems, predatory contact, commercial profiling, and social pressures that are difficult for adults to fully see, let alone regulate. Governments are right to take these risks seriously. Parents are right to be concerned. Schools are right to see the consequences spilling into classrooms, friendships, wellbeing, and behaviour.

But a real problem does not make every proposed solution a good one.

The central flaw in the current policy direction is that it treats access control as though it were the same thing as safety. If children can be kept out, the thinking goes, then children can be protected. But the evidence and arguments considered throughout this essay point in a different direction. Age-verification systems are easy to evade, difficult to enforce evenly, and structurally dependent on trade-offs between effectiveness and intrusiveness. The weakest systems are easily bypassed. The strongest systems create new privacy, security and identity risks. Between those poles sits a wide range of imperfect methods that may provide friction, but not certainty.

The social media ban reveals the problem clearly. It is framed as a child-safety measure, but its practical effect depends on platforms being able to reliably distinguish children from adults, and on young people not adapting around the system. That was never a safe assumption. Workarounds are simple, shared, and socially normalised. Some users will lie. Some will borrow accounts. Some will use VPNs. Some will move to less visible spaces. A policy that can be bypassed so easily should not be treated as a stable foundation for broader regulation.

The proposed Children's Online Privacy Code raises the stakes. Unlike the Social Media Ban, it is not presented as an exclusionary measure. It is framed as a privacy protection, and much of its purpose is legitimate. Children's personal information should be handled with greater care. Services should not be free to datafy childhood, manipulate consent, or design online environments that exploit vulnerability. But the mechanism matters. If the Code encourages more services to ascertain age in order to decide which obligations apply, then it risks extending age-assurance logic across much wider areas of online life.

That is the danger at the heart of this policy trajectory. A system built to protect children may end up asking everyone to prove, declare, estimate, infer, or verify who they are. It

may normalise age classification as a condition of access. It may encourage more services to collect or process sensitive information. It may create new markets for verification providers, new incentives for behavioural inference, and new stores or flows of identity data. It may also leave the least compliant, least accountable spaces relatively untouched.

This is not a hypothetical concern. Australia has already seen what happens when identity datasets are compromised. Driver's licence numbers, passport numbers, Medicare details, dates of birth, addresses, and sensitive personal information have all been exposed in major breaches. A password can be changed. A compromised identity document is much harder to repair. A face, date of birth, or history of identity exposure cannot simply be recalled. Any policy that increases the routine production of breach-worthy data should be treated with caution, not reassurance.

Nor can the problem be solved by pointing to future privacy-preserving technology. Better systems may be possible. Some may eventually allow age eligibility to be proven without exposing full identity. Those developments should be encouraged. The choice is not between inaction and fragile gates. A better policy would focus on the systems that create harm. It would raise the privacy floor for everyone, not build a larger machinery for sorting children from adults. It would support schools, parents, carers, and young people directly. It would keep age assurance narrow, necessary, proportionate, independently scrutinised, and genuinely privacy-preserving where it is unavoidable.

Child protection policy should be judged not by how serious it sounds, but by what it actually produces. If it produces routine circumvention, uneven enforcement, displacement into worse spaces, and broader identity exposure, then it has not solved the problem. It has merely changed its shape. Worse, it may make future solutions harder by normalising infrastructure that should never have been built at scale without clearer limits.

A safer internet cannot be built on fragile gates. It cannot depend on checkpoints that determined users can step around while ordinary users are asked to surrender more information. It cannot protect children by making everyone more identifiable, more trackable, and more exposed. If Australia is serious about protecting young people online, it must move beyond the illusion of control and confront the design, data, and accountability failures that make digital environments unsafe in the first place.

The goal should not be an internet where every door demands proof. The goal should be an internet where fewer doors lead to harm.

Reference list

Alajaji, R 2025, *10 (Not So) Hidden Dangers of Age Verification*, Electronic Frontier Foundation.

Allen, T, McColl, L, Walters, K & Evans, H 2022, *Measurement of Age Assurance Technologies*.

Australian Government 2024, *Federal Register of Legislation - Online Safety Amendment (Social Media Minimum Age) Act 2024*, Legislation.gov.au.

Barrett, J 2023, 'Latitude Financial cyber-attack worse than first thought with 14m customer records stolen', *The Guardian*.

Bellovin, S 2025, *Privacy-Preserving Age Verification-and Its Limitations*.

Bernstein, M, Monroy-Hernández, A, Harry, D, André, P, Panovich, K & Vargas, G 2011, *4chan and /b/: An Analysis of Anonymity and Ephemerality in a Large Online Community*, MIT.

Croft, D 2025, *80% of under-13s bypass social media age restrictions*, Cyberdaily.au, Cyber Daily.

Crowley, T 2024, 'Australian-first sanctions target Russian accused of Medibank hack', *ABC News*, 22 January.

Cubitt, T & Morgan, A 2024, *Trends & issues in crime and criminal justice*, Australian Institute of Criminology.

Department of Infrastructure, Transport, Regional Development, Communications, Sport and the Arts 2025, *Age Assurance Technology Trial— Final Report*, Department of Infrastructure, Transport, Regional Development, Communications, Sport and the Arts.

Dervisevic, H 2025, *Age verification errors see some under-16s retain access to banned social media platforms*, Abc.net.au, ABC News.

Eltaher, F, Gajula, R, Miralles-Pechuán, L, Thorpe, C & Mckeever, S 2025, *The Digital Loophole: Evaluating the Effectiveness of Child Age Verification Methods on Social Media*.

eSafety Commissioner 2022, *Which platforms are age-restricted? | eSafety Commissioner*, eSafety Commissioner.

— 2026, *Social Media Minimum Age compliance report*, eSafety Commissioner.

Gazy, M 2026, *'Left out and ostracised': The social media ban's uneven execution — and other effects*, SBS News.

Gilbert, F 2008, *Age Verification as a Shield for Minors on the Internet: A Quixotic Search?*, UW Law Digital Commons.

Kaye, B 2025, 'Australia leader defends social media ban as teens brag about staying online', *Reuters*, 11 December.

Ling, J 2023, *Inside 4chan's top-secret moderation machine*, WIRED, viewed 30 April 2026, <https://www.wired.com/story/4chan-moderation-buffalo-shooting/>.

Lueks, W, Dreyer, S, Federrath, H & Simon, J 2026, 'Assessing Age Assurance Technologies: Effectiveness, Side-Effects, and Acceptance', *arXiv (Cornell University)*, Cornell University.

Office of the Australian Information Commissioner 2025, *Latest Notifiable Data Breach statistics for January to June 2025*, OAIC.

— 2026a, *Children's Online Privacy Code*, OAIC.

— 2026b, *Exposure Draft Explanatory Statement*, 31 March, OAIC.

Queensland Government 2022, *Optus data breach*, Queensland Government.

Randall, A & Guest, A 2026, 'First figures released for social media ban - ABC listen', *ABC listen*, Australian Broadcasting Corporation, viewed 29 April 2026, <http://abc.net.au/listen/programs/worldtoday/first-figures-released-for-social-media-ban/106237464>.

Rescorla, E, Arnao, Z & Cooper, A 2026, *Age Assurance Online: A Technical Assessment of Current Systems and Their Limitations*.

Sas, M & Mühlberg, JT 2024, *Trustworthy age assurance?*, Greens/EFA.

Taylor, J 2025, *4chan unlikely to be included in Australia's under-16s social media ban, eSafety commissioner says*, the Guardian, The Guardian.

Thorleifsson, C 2021, 'From cyberfascism to terrorism: On 4chan/pol/ culture and the transnational production of memetic violence', *Nations and Nationalism*, vol. 28, no. 1, pp. 286–301.

UNICEF 2025, *Age restrictions alone won't keep children safe online*, Unicef.org.

Unforeseen Consequences

How Child Safety Policy Risks Expanding Identity Exposure

Executive Summary

A concise summary of the policy critique on Australia's proposed Children's Online Privacy Code, social media age verification, and the risks of expanding age-assurance infrastructure.

Purpose and Central Argument

This submission responds to Australia's current policy direction on children's online safety and privacy, with particular focus on the proposed *Privacy (Children's Online Privacy) Code 2026* and its relationship to broader age-verification and age-assurance measures. It supports stronger protection for children online. Children deserve digital environments that are safer, more transparent, less exploitative, and less dependent on the collection and commercial use of personal information.

The proposed Code contains important privacy objectives, including stronger privacy defaults, clearer information for children, tighter consent requirements, and greater control over personal information (OAIC, 2026a; OAIC, 2026b). However, the mechanism used to pursue those aims requires closer scrutiny. Section 8 may pressure covered services to ascertain whether users are children before collecting personal information, unless those services apply the Code's protections to all users regardless of age (OAIC, 2026a). In practice, this may encourage broader use of age-assurance systems across digital services "likely to be accessed by children" (OAIC, 2026a; OAIC, 2026b).

This submission proceeds from a simple premise: the problem is real; the current mechanism is inadequate; and expanding that mechanism before its failures are addressed risks broader harm. Age verification and age assurance are often easy to bypass, unevenly enforceable, and most effective against compliant users and services rather than determined users or non-compliant platforms (Gilbert, 2008; Lueks et al., 2026). Stronger forms of age assurance can also require intrusive identity checks, biometric estimation, behavioural profiling, parental attestation, or third-party verification (Allen et al., 2022; Rescorla, Arnao & Cooper, 2026). The result is a policy pathway that may fail to reliably protect children while normalising age classification, identity exposure, and new cybersecurity risks across ordinary online life.

Key Risks

The concern is not that age assurance may fail in one way, but that several weaknesses may operate together: children can work around the gates, compliant services carry the burden, stronger checks create privacy risks, and some harms may move out of sight.

A Gate Children Can Walk Around

Age verification is often presented as a boundary between children and harmful online spaces. In practice, it is more likely to create friction than certainty. Young people can bypass age gates by entering false dates of birth, using existing accounts, borrowing devices, relying on adult-assisted access, using VPNs, or moving to services that do not apply Australian restrictions. These workarounds need only be common enough to break the connection between the person verified and the person using the service.

Age checks may still delay or deter some children, particularly younger children. But a policy that can be routinely bypassed should not be treated as a reliable child-safety mechanism. Its most likely outcome is uneven friction: compliant users and services face new burdens, while determined users learn how to move around them (Gilbert, 2008; Lueks et al., 2026).

From Social Media Ban to System-Wide Age Checks

The concern does not stop with social media age restrictions. The proposed Code may extend the same logic into a broader privacy framework. The Code applies to certain online services that are "likely to be accessed by children", and section 8 requires covered entities to take reasonable steps to ascertain an end-user's age before collecting personal information, unless the entity applies the Code's protections to all users regardless of age (OAIC, 2026a).

That exception is important, and the Code is not written as a universal identity law. However, it may still create practical pressure for services to classify users by age. Many general-audience services are likely to be accessed by children, even if they are not designed for them. Faced with uncertainty, services may choose the safest compliance path: ask users to declare, prove, estimate, or otherwise establish age before access or data collection. In this way, a privacy reform could unintentionally normalise age assessment across ordinary online life (OAIC, 2026a; OAIC, 2026b).

The Age-Assurance Trade-Off

"Age assurance" is not one method. It is a broad category that includes self-declaration, age inference, age estimation, age verification, and parental attestation (OAIC, 2026b). These methods sit on a ladder. At the bottom are light-touch methods that collect little information but are easy to bypass. At the top are stronger methods that may be harder to evade, but are more intrusive.

Self-declaration, such as entering a date of birth, is cheap and privacy-light, but weak. Behavioural age inference may avoid asking for identity documents, but it depends on profiling users through account activity, browsing behaviour, platform signals, or other data traces. Biometric or quasi-biometric age estimation may use a face, voice, or other bodily signal. Document-based verification may provide stronger assurance, but can expose far more information than age alone.

This is the central unresolved trade-off: the less intrusive the system is, the easier it is to bypass; the harder it is to bypass, the more intrusive it tends to become. That trade-off should sit at the centre of any decision to expand age-assurance expectations (Allen et al., 2022; Lueks et al., 2026; Rescorla, Arnao & Cooper, 2026).

When Verification Data Becomes Breach Data

The privacy risk becomes most serious when age assurance relies on sensitive information. A weak age gate may fail. A strong age gate may create breach-worthy data. Identity documents, facial images, biometric estimates, account signals, parental verification, and third-party checks can all create new points where personal information is collected, transmitted, inferred, logged, retained, or mishandled.

This risk is not theoretical. The Optus, Latitude Financial, and Medibank breaches are only the most visible examples of a broader problem: the OAIC received 532 notifiable data breach reports in the first half of 2025 alone (OAIC, 2025). These incidents show that even well-resourced Australian organisations can fail to protect sensitive personal information once collected. Age-assurance systems would not necessarily collect the same information, but the lesson is clear: once identity material enters a system, the public must rely on its security, retention rules, vendors, governance, and future behaviour.

Privacy-preserving age assurance may reduce these risks, but it cannot be assumed. Stronger designs, such as selective disclosure or zero-knowledge proofs, require careful implementation and are not yet the ordinary reality of age assurance. Until such safeguards are mature, enforceable, and independently audited, policymakers should be cautious about encouraging more services to generate verification data in the first place (Queensland Government, 2022; Barrett, 2023; Crowley, 2024; Sas & Mühlberg, 2024; Rescorla, Arnao & Cooper, 2026).

Moving Harm Out of Sight

Age restrictions may also create a displacement problem. If young people are blocked, removed, or discouraged from using mainstream platforms, it does not follow that they simply go offline. Some may create new accounts, use workarounds, or move into less visible spaces such as anonymous forums, private servers, encrypted groups, fringe platforms, or offshore services.

Mainstream platforms, while deeply flawed, are at least partly visible and accountable: parents know their names, teachers hear about conflicts that begin there, regulators can contact them, researchers and journalists can study them, and users can report harmful content. Less visible spaces may have weaker moderation, fewer reporting pathways, more anonymity, stronger anti-regulatory cultures, and greater exposure to harmful content or communities.

A reduction in visible underage accounts is not the same as a reduction in harm. If young people are pushed from imperfectly regulated spaces into barely regulated ones, the policy may make harm harder to see and interrupt. Child-safety policy should reduce the conditions that produce harm, not simply move risk out of view (eSafety Commissioner, 2022; Cubitt & Morgan, 2024; UNICEF, 2025).

What Better Policy Would Look Like

A better approach would protect children by making online environments safer, not by making every user more identifiable.

Raise the privacy floor for everyone. Where the Code's protections would benefit the wider public, they should not depend on identifying who is a child. Limits on unnecessary data collection, stronger privacy defaults, clearer consent, restrictions on profiling, and better rights over personal information would benefit children and adults alike.

Make age assurance exceptional. Age assurance should be used only where necessary, proportionate, and clearly justified by evidence. It should not become a general compliance shortcut for any service that might be accessed by a child.

Require strict data minimisation and independent audit. Where age assurance is genuinely necessary, systems should collect the minimum information required, avoid retaining identity documents or biometric data, and prohibit reuse of verification data for advertising, profiling, analytics, or unrelated commercial purposes. Third-party providers should be independently audited and subject to strong breach, security, and deletion obligations.

Protect anonymous and pseudonymous access. Privacy reform should not casually make identification the default condition of participation. Users should retain the ability to browse, seek information, make complaints, access support, and participate pseudonymously where identification is not genuinely necessary.

Regulate platform design directly. Many online harms are produced by platform systems themselves: recommender algorithms, manipulative consent flows, addictive engagement loops, dark patterns, profiling, weak reporting systems, unsafe defaults, and poor moderation. Child safety should not be reduced to who gets through the gate.

Conclusion: A Safer Internet Cannot Be Built on Fragile Gates

Children deserve meaningful protection online, and platforms should be held accountable for the environments they create. But protection should not depend on making Australians more identifiable, traceable, or easily classified. Age assurance may have a narrow role in genuinely high-risk contexts, but it should not become the foundation of children's privacy policy.

The current trajectory risks building a system that determined users can work around, while ordinary users and compliant services absorb the privacy burden. A safer internet requires online services to collect less, profile less, manipulate less, and take responsibility for the risks they create.

Reference list

Allen, T, McColl, L, Walters, K & Evans, H 2022, *Measurement of Age Assurance Technologies*.

Barrett, J 2023, 'Latitude Financial cyber-attack worse than first thought with 14m customer records stolen', *The Guardian*.

Crowley, T 2024, 'Australian-first sanctions target Russian accused of Medibank hack', *ABC News*, 22 January.

Cubitt, T & Morgan, A 2024, *Trends & issues in crime and criminal justice*, Australian Institute of Criminology.

eSafety Commissioner 2022, *Which platforms are age-restricted? | eSafety Commissioner*, eSafety Commissioner.

Gilbert, F 2008, *Age Verification as a Shield for Minors on the Internet: A Quixotic Search?*, UW Law Digital Commons.

Lueks, W, Dreyer, S, Federrath, H & Simon, J 2026, 'Assessing Age Assurance Technologies: Effectiveness, Side-Effects, and Acceptance', *arXiv (Cornell University)*, Cornell University.

Office of the Australian Information Commissioner 2025, *Latest Notifiable Data Breach statistics for January to June 2025*, OAIC.

— 2026a, *Children's Online Privacy Code*, OAIC.

— 2026b, *Exposure Draft Explanatory Statement*, 31 March, OAIC.

Rescorla, E, Arnao, Z & Cooper, A 2026, *Age Assurance Online: A Technical Assessment of Current Systems and Their Limitations*.

Sas, M & Mühlberg, JT 2024, *Trustworthy age assurance?*, Greens/EFA.

UNICEF 2025, *Age restrictions alone won't keep children safe online*, Unicef.org.