Executive Summary

Away from Keyboard (AFK) Inc. is an Australian charity founded in 2023 to protect children and young people from online harm and to empower families, carers and communities to navigate the digital world safely. AFK delivers digital safety education, develops Safety by Design frameworks, and advocates globally for children's rights online through partnerships with the United Nations, international networks, and Australian regulators.

My commitment to this work is deeply personal. As a teenager, I was groomed online by an older man and coerced into a forced marriage. That experience exposed me to exploitation and harm at a formative stage of my life. It showed me first-hand how easily technology can be weaponised to manipulate vulnerable young people and evade oversight. I founded Away from Keyboard to turn that painful experience into action, to ensure that no child endures what I did, and that technology is used to protect rather than exploit.

AFK's mission and my personal expertise align directly with the Committee's inquiry. We work at the intersection of policy, technology and community-level intervention, running workshops, advising on policy, and building practical tools to ensure that Safety by Design becomes the basic standard rather than an aspiration.

This submission highlights three systemic challenges:

Privacy-preserving and equitable age assurance, verification methods must not create new risks for children, carers or older Australians.

Digital literacy and resilience at the community scale, lifting capacity across parents, carers, and young people so that technology restrictions translate into real-world protection.

Closing the gaps left by current codes, including "logged out" exposure to harmful but lawful content, scams, and technology-facilitated gender-based violence, especially as generative AI accelerates new forms of harm.

About Us

Away from Keyboard (AFK) Inc. was founded in 2023 in regional Victoria to prevent harm to children and young people online by combining lived experience, community education, and high-level policy advocacy. AFK's work is anchored in the UN Convention on the Rights of the Child, Safety by Design principles, and trauma-informed practice. Our model is proactive, not reactive: we equip children, carers, educators and communities with the knowledge, skills and tools to identify and avoid online risks before exploitation or abuse occurs.

My lived experience of being groomed online and coerced into a forced marriage as a teenager drives this preventative approach. AFK exists to ensure other children do not endure similar harm and that technology is built and governed to protect rather than exploit.

In addition to leading AFK, I hold multiple advisory roles with the National Council of Women Victoria (NCWV), including Adviser for Rural & Regional Women, Youth, ICT & Mass Media, and Human Rights, and I serve as Chief Revenue Officer for Women 4 STEM, a national not-for-profit supporting the entry, retention and advancement of women in Science, Technology, Engineering and Mathematics. This gives me a unique, cross-sector perspective on how online safety, ethical technology and emerging risks affect children, families and future workers. AFK and I have also written to the Prime Minister and Federal Ministers advocating for stronger online safety measures, explicitly responding to the UN Special Rapporteur's report on the sale and sexual exploitation of children. providing me with a unique, cross-sector perspective on how online safety policies impact different communities.

Key Achievements and Contributions

International Leadership on Prevention: Delivered Unsafe by Design: The Unseen Risks of AI for the Girl Child at the United Nations in New York, framing AI safety and technology-facilitated gender-based violence as a preventable harm issue. Contributed to Geneva Peace Week on human rights due diligence in emerging tech, advancing global norms on child-safety-by-design. AFK has also endorsed the Global Digital Compact, Osaka Protocol, and the Hamburg Declaration.

Policy and Advocacy for Prevention: As Adviser with NCWV, provided evidence-based recommendations on online safety, youth protection and rural/regional digital inclusion. Drafted and co-signed letters to the Prime Minister advocating stronger online safety responses and implementing Safety by Design principles in national legislation.

Community-Level Harm Prevention: Designed and delivered workshops across regional and remote communities, building carers' and young people's capacity to identify grooming, scams, and coercive control online before harm occurs.

Resources for Low-Literacy Carers: Created plain-language toolkits and step-by-step guides on privacy, scams and age assurance, enabling vulnerable adults to protect children in their care.

Recognition of Leadership: Recognised by the Australian Women in Security Awards as Best Volunteer for my work in digital child safety. Finalist in the Women in AI – APAC Social Good category for my advocacy on ethical AI and community protection. These honours reflect my sustained national and international leadership in online safety and ethical technology.

Thought Leadership on Emerging Technology: Invest significant time speaking and writing about the dangers of emerging technologies, ethical AI, and the impact of algorithmic and platform design on children, families and communities.

AFK's work sits at the front line of harm prevention. We identify gaps in regulation, build public capability, and push for system-level changes that reduce risk at its source. .

Response to Terms of Reference

1. Privacy and Data Protection Implications of Age Verification

Age-verification systems can reduce children's exposure to harm, but if they are built on identity capture or biometrics they risk creating a new market in children's and carers' most sensitive data. This risk is amplified for older Australians and regional communities who often have lower digital literacy and less access to trusted support, making them more likely to hand over unnecessary information or abandon services entirely.

Globally, about one in eight children experienced the non-consensual taking, sharing, or exposure to sexual images in the past year, and 12.5% experienced online solicitation (Lu et al. 2024)[1]. This underscores the need for effective protection but also for strict privacy safeguards when deploying age checks. The eSafety Commissioner's BOSE periodic notices indicate gaps in transparency, with portions of provider responses withheld from publication, highlighting the ongoing challenge of seeing how platforms actually handle sensitive data (eSafety Commissioner 2025)[2]. Australian research shows a pronounced digital divide: parents and carers, especially those aged 55+, frequently rate their children's tech understanding as better than their own (Engelmann et al. 2025)[3]. This imbalance increases the risk of mistakes when navigating age-assurance flows or privacy settings, and it is often steeper in communities with fewer local programs.

Young people themselves recommend privacy-enhancing age verification, including zero-knowledge proofs, reusable age tokens, and third-party age-verification services to minimise data exposure and keep sensitive ID off the platform (Normative Advances on TF VAWG 2024)[4]. They also warn about bias and discrimination when AI/ML is used in age checks and call for methods that work for all users regardless of identity documents or digital skills (Normative Advances on TF VAWG 2024)[4].

Recommendations

A. Put children's rights and privacy first

- Require a Child Rights and Privacy Impact Assessment (CRPIA) for every age-assurance deployment, independently reviewed and published.
- Codify data minimisation, purpose limitation, non-reidentification and strict deletion timelines for any age-signal or metadata associated with minors and their carers, modelled on GDPR and Lanzarote principles.

B. Prevent function-creep and create accountability

- Ban any secondary use of age-assurance data for targeting, profiling, personalisation, or monetisation.
- Create a publicly searchable Register of Age-Assurance Providers, with annual plain-English audit summaries and complaint outcomes.

C. Make verification privacy-preserving by design

- Prefer tokenised age assertions and zero-knowledge proofs so platforms receive only a yes/no age attribute and never hold raw identity data [4].
- Require method plurality: no single biometric route. Vendors must provide non-biometric options and clearly explain trade-offs, bias controls, and error handling [4].

D. Design for older Australians and regional communities

- Fund assisted and offline pathways to complete age-assurance in trusted places (libraries, councils, community centres) and require vendors to deliver plain-language, step-by-step materials.
- Mandate usability and accessibility testing with low-literacy and 55+ cohorts, and publish the results [3].
- Require clear redress: human help channels, transparent error messages, and simple dispute processes when users fail an age check.

Measures of Success

- 100% of vendors listed on the public register; all CRPIAs published with clear mitigations and deletion schedules.
- Independent usability studies show ≥90% task completion for 55+ and low-literacy users without external help; drop-off rates below a set threshold.
- Verified deletion of age-assurance artefacts within defined time windows.
- Year-on-year increase in carers' self-reported confidence to manage online safety and verification tasks in regional/rural areas.

References

[1] Lu, M. et al. (2024). *A Content Analysis of Metrics on Online Child Sexual Exploitation and Abuse Used by Online Content-Sharing Services*.

[2] eSafety Commissioner. (2025). *Basic Online Safety Expectations (BOSE) Periodic Notice – CSEA & Sexual Extortion*.

[3] Engelmann, L. et al. (2025). *Developing Quality Standards for Community-Based Online Child Sexual Exploitation and Abuse Interventions*.

[4] UN Women. (2024). *Normative Advances on Technology-Facilitated Violence Against Women and Girls*.

2. Expansion of Corporate Data Collection and Profiling Capabilities

The Online Safety Code could unintentionally entrench surveillance-based business models. Without tight regulation, age-assurance data and behavioural analytics may be collected far beyond what is necessary to establish age, creating permanent profiles of children, carers and older Australians. These risks are acute for regional and low-literacy communities, where users may be less able to scrutinise privacy policies or exercise their rights.
We already know that scam networks exploit personal data: in Southeast Asia, 63% of adults reported being scammed in the last 12 months and 22% lost money (GASA 2023)[1]. Profiling minors and carers, or exposing their verification data to third parties, would give scammers and hostile actors even more high-value targets.

The UN Normative Advances on Technology-Facilitated Violence Against Women and Girls (TF VAWG) explicitly call for human rights due diligence in online services, noting that women, girls and marginalised users are disproportionately harmed by opaque data practices (UN Women 2024)[2].

Risks to watch

1. Function creep, age-assurance vendors using identity data for profiling, ads or biometric research.
2. Data brokers, cross-platform linking of age data to existing behavioural datasets, enabling unprecedented microtargeting of children and their carers.
3. Discrimination, AI systems inferring sensitive characteristics (disability, ethnicity, family structure) from behavioural patterns, then delivering biased experiences or pricing.

Recommendations

A. Create a National Register of Age-Assurance Providers

- Mandatory public reporting on data handling, retention, breach incidents, and privacy safeguards.
- Annual independent audits with plain-English summaries available to the public.

B. Ban Secondary Use and Profiling of Verification Data

- Enshrine in the Code that any data collected to establish age cannot be reused for targeting, profiling or personalisation.

- Explicitly prohibit selling or sharing age-assurance data with data brokers, marketers, or unrelated government programs.

C. Require Privacy-by-Default and Data Minimisation

- All services likely to be accessed by minors must be privacy-by-default, with only minimal data collection strictly necessary to provide the service.
- Explicitly prohibit behavioural advertising to under-18 users.

D. Extend Protection to Carers and Older Australians

- Recognise that many older Australians manage online accounts on behalf of children. Require age-assurance flows to protect carers' data with the same safeguards as minors.
- Fund community-based awareness programs on privacy risks, particularly in regional and remote areas.

Measures of Success

- All age-assurance vendors listed on the National Register, with independent audits published annually.
- Zero confirmed cases of secondary use of verification data.
- Year-on-year decline in minors and carers included in ad-tech datasets.
- Rising digital literacy among regional and older Australians measured by ACCCE-style baseline surveys.

Evidence

- Global Anti-Scam Alliance. (2023). *State of Scams Southeast Asia*, 63% of adults scammed in 12 months; 22% lost money [1].
- UN Women. (2024). *Normative Advances on Technology-Facilitated Violence Against Women and Girls*, calls for human rights due diligence [2].

References

[1] Global Anti-Scam Alliance. (2023). *State of Scams Southeast Asia*.
[2] UN Women. (2024). *Normative Advances on Technology-Facilitated Violence Against Women and Girls*.

3. Technical Implementation and Efficacy of Age Verification and Content Filtering

Age gates and filters alone cannot stop grooming, livestream abuse, or AI-generated child sexual exploitation material (CSAM). Children can still bypass restrictions using VPNs, borrowed accounts, or "logged-out" browsing, a space where harmful but legal ("lawful but awful") content remains freely accessible. Unless the Code addresses these loopholes, much of the risk will simply migrate to channels that are harder to monitor.

Generative AI further intensifies this challenge. The eSafety Commissioner's BOSE Periodic Notice found no provider fully proactive on livestreaming CSEA and documented a 1,325% increase in generative-AI CSAM year-on-year [1]. WeProtect

Global Alliance reported a 360% rise in self-generated sexual images of 7–10-year-olds, underscoring how quickly new harms emerge when detection and filtering lag [2].

Key gaps

- Logged-out loophole: Filters typically apply to logged-in accounts; search results, link previews and "incognito" browsing remain largely unfiltered.
- VPN circumvention: Age gates based on IP or device ID are easily bypassed.
- Static filtering: Keyword filters cannot keep pace with generative AI, new slang, or new types of imagery.
- No systematic red-team testing: Platforms self-attest to efficacy without independent verification.

Recommendations

A. Make Safety by Design Mandatory for All Access Points

- Require SafeSearch defaults, blurred previews, TFGBV filters and crisis-card interventions even for non-logged-in traffic.
- Extend filtering and risk controls to autocomplete, recommendation engines, and link previews.
- Mandate a minimum set of friction measures for high-risk categories, e.g. confirm age before repeated access to violent or sexual content.

B. Independent Red-Team Testing and Harm Simulations

- Mandate third-party red-team testing of age-assurance and filtering systems before rollout and at regular intervals.
- Include AI-based harm simulations to identify new pathways (e.g., generative AI or new encryption methods).

C. AI-Assisted Grooming and Sextortion Detection

- Require platforms to deploy privacy-preserving AI models trained to detect grooming patterns, sextortion, and child sexual exploitation attempts, with strict guardrails against false positives and misuse.
- Ensure survivor-led training for moderators to contextualise flagged content.

D. Usability and Accessibility Across Communities

- Fund regional pilots to test whether filters and controls work under slower connections, older devices, and low digital literacy conditions.

- Publish results and iterate design with real users, including children, carers and older Australians.

Measures of Success

- Year-on-year reduction in time-to-detection for grooming and sextortion content.
- Percentage of harmful content blurred or blocked for non-logged-in under-16 traffic.
- Documented improvement in user experience across regional and low-literacy cohorts.
- Public "filter efficacy" dashboards showing detection rates, bypass attempts prevented, and generative AI content flagged.

Evidence

- eSafety Commissioner. (2025). *Basic Online Safety Expectations (BOSE) Periodic Notice – CSEA & Sexual Extortion*, no provider fully proactive on livestreaming CSEA; 1,325% increase in generative-AI CSAM [1].
- WeProtect Global Alliance. (2023). *Global Threat Assessment*, 360% rise in self-generated sexual images of 7–10-year-olds [2].

References

[1] eSafety Commissioner. (2025). *Basic Online Safety Expectations (BOSE) Periodic Notice – CSEA & Sexual Extortion*.

[2] WeProtect Global Alliance. (2023). *Global Threat Assessment*.

4. Alternative Technical Approaches to Online Safety for All Users

Current age-assurance and filtering systems assume every parent or carer can manage complex settings, yet Australia lacks national infrastructure to support low-literacy adults, carers and regional families. The risk is that regulation becomes a compliance exercise for industry but leaves communities without the skills or tools to apply it effectively.

Digital literacy is the linchpin of effective online safety. Without it, parents and carers cannot support children through age verification, privacy settings or scam prevention.

This gap is especially acute in regional and remote areas, where library closures, limited public transport and slower connectivity compound the problem.

The Australian Centre to Counter Child Exploitation (ACCCE) research shows low public awareness and underfunded community interventions, despite the scale of online exploitation harms (Engelmann et al. 2025)[1]. Meanwhile, the Global Anti-Scam Alliance found 63% of adults in Southeast Asia were scammed in 12 months, and 22% lost money, showing how low-literacy communities are disproportionately affected (GASA 2023)[2].

Key gaps

- No national digital literacy program dedicated to online safety and age-assurance.
- Fragmented parental-control tools with no interoperability across devices and platforms.
- Youth excluded from design of safety measures despite being the primary users affected.

Recommendations

A. Establish a National Digital Literacy & Age-Assurance Support Program

- Provide ring-fenced funding for libraries, councils, NGOs and schools to deliver face-to-face support, particularly in regional and remote communities.
- Develop plain-language, multilingual curricula covering age-assurance, scam recognition, privacy settings and reporting harmful content.
- Offer micro-credentials and train-the-trainer programs to build local capacity.

B. Mandate Open Parental-Control and Safety APIs

- Require all large online services and device manufacturers to provide open, standards-based parental-control and safety APIs.
- Allow third-party innovators to build accessible dashboards and cross-device controls for families.
- Create an Australian Safety-Tech Standards Body to govern interoperability.

C. Pilot a Youth Safety Innovation Lab

- Fund a national Youth Safety Innovation Lab bringing together young people, survivors, tech providers and researchers to co-design safer UX and test emerging tools.

- Use the Lab to red-team new features, evaluate age-assurance methods and generate youth-led solutions.

D. Embed Safety by Design Beyond Compliance

- Link grant funding and procurement to demonstrable Safety by Design outcomes.
- Offer incentives to platforms that release open-sourced safety tools, share best practices and fund independent evaluation.

Measures of Success

- Percentage of regional households with access to in-person digital literacy support.
- Uptake rates of open parental-control APIs by Australian developers.
- Youth participation rates in safety co-design processes.
- Documented improvements in carers' and young people's confidence managing online safety settings.

Evidence

- Engelmann, L. et al. (2025). *Developing Quality Standards for Community-Based Online Child Sexual Exploitation and Abuse Interventions*, low public awareness and underfunded community interventions [1].
- Global Anti-Scam Alliance. (2023). *State of Scams Southeast Asia*, widespread scam exposure [2].

References

[1] Engelmann, L. et al. (2025). *Developing Quality Standards for Community-Based Online Child Sexual Exploitation and Abuse Interventions*.

[2] Global Anti-Scam Alliance. (2023). *State of Scams Southeast Asia*.

5. Appropriate Oversight Mechanisms for Online Safety Codes

Oversight of online safety codes in Australia currently emphasises process compliance rather than outcome effectiveness. Platforms can self-report on their adherence to codes, but there is no systematic way to verify whether children are actually safer, harmful content is actually reduced, or carers have become more digitally literate. Without rigorous oversight, regulation risks becoming symbolic rather than transformative.

Global best practice recognises that child rights, privacy, and safety outcomes must be measurable. The UN Special Rapporteur on the Sale and Sexual Exploitation of Children calls for harmonised child-rights data collection, independent oversight mechanisms, and direct involvement of young people in shaping policy and evaluating implementation (UN Human Rights Council 2025)[1].

Key gaps

- Opaque self-assessments: Providers publish few metrics or use inconsistent definitions.
- No independent verification of harm reduction: We do not track outcomes such as reduction in grooming attempts, sextortion, or scam losses.
- Youth perspectives underrepresented: Policies are designed without structured youth input, risking misalignment with real-world needs.

Recommendations

A. Establish a Standing Youth Advisory Council under the Online Safety Act

- Bring together diverse young people (including from regional and marginalised communities) to test communications, UX and harm mitigation measures.
- Provide stipends and training to enable authentic participation rather than token consultation.

B. Require Annual Public Dashboards Reporting on Harm Reduction

- Mandate that platforms and the regulator publish consistent, age-disaggregated metrics on key harms: grooming attempts, CSAM removal, scam incidents, and parental-control uptake.
- Include a measure of digital literacy uplift among carers and children, linked to community-based interventions.

C. Tie Industry Compliance to Measurable Improvements, Not Only Box-Ticking

- Define clear outcome indicators (e.g. median time to remove harmful content, reduction in logged-out exposure, improvement in safe default settings).
- Apply a “comply or explain” regime where failure to meet benchmarks triggers corrective action or penalties.

D. Independent Evaluation of the Codes

- Commission independent academic and NGO-led evaluations of the Codes every two years, with full public release of findings.
- Integrate survivor and youth-led audits to ensure policies reflect lived experience.

Measures of Success

- Annual dashboards showing year-on-year reduction in harmful exposures and improvements in digital literacy metrics.
- Structured youth feedback incorporated into code revisions and published in plain English.
- Verified compliance audits replacing self-attestation as the primary oversight mechanism.
- Transparent public data enabling researchers, NGOs and media to scrutinise performance.

Evidence

- UN Human Rights Council. (2025). *Report of the Special Rapporteur on the Sale and Sexual Exploitation of Children, A/HRC/58/52/Add.1*, calls for harmonised child-rights data collection and oversight mechanisms [1].
- International examples (UK Age Appropriate Design Code, EU Digital Services Act) show that public dashboards and independent regulators improve trust and compliance.

References

[1] UN Human Rights Council. (2025). *Report of the Special Rapporteur on the Sale and Sexual Exploitation of Children, A/HRC/58/52/Add.1*.

6. Global Experience and Best Practice

Australia's online safety framework is robust by international standards but risks falling behind as other jurisdictions embed Safety by Design, child rights impact assessments, and algorithmic accountability into law. Aligning with global norms enables consistency for multinational platforms, accelerates compliance, and raises protections for Australian children. It also prevents the "jurisdiction shopping" where harmful services relocate to the least regulated environment.
UN CRC General Comment No. 25 explicitly calls for governments to ensure that digital services respect children's rights to privacy, safety and participation from the design stage onwards. The UK Age Appropriate Design Code and EU Digital Services Act provide models for risk assessments, data minimisation, and accountability dashboards. The

Council of Europe Lanzarote Convention and the Hamburg Declaration both emphasise international harmonisation and Safety by Design principles (UN Women 2024; Hamburg Declaration 2023)[1][2].

Key gaps

- Fragmentation: Without harmonisation, platforms may apply lower standards in Australia than overseas.
- Narrow focus: Current codes focus heavily on age and content but less on algorithmic design, recommender systems, and cross-border data flows.
- Limited international cooperation: Few formal agreements on data sharing, enforcement, or youth-participation benchmarks.

Recommendations

A. Align Code Provisions with Global Child Rights Standards

- Incorporate UN CRC General Comment No. 25 into the Online Safety Code as a baseline.
- Adopt UK-style "age-appropriate design" standards for default settings, privacy, and transparency.

B. Integrate the 4Cs Risk Model Across All Guidance and Audits

- Systematically assess Contact, Content, Conduct, and Contract risks to ensure a holistic approach.
- Require providers to show how they address each "C" with measurable outcomes.

C. Ratify and Implement the Council of Europe Lanzarote Convention

- This treaty strengthens cross-border cooperation on child protection online, aligning Australia with European partners and global best practice.

D. Build International Partnerships on Enforcement and Research

- Establish formal data-sharing agreements with other regulators to accelerate removal of harmful content across borders.
- Partner with global youth-led organisations to pilot safety innovations and embed international peer review.

E. Incentivise Safety Innovation

- Offer recognition or procurement preference to companies that voluntarily exceed international standards (e.g., publish algorithmic impact assessments, open-source safety tools).

Measures of Success

- Formal adoption of international child-safety frameworks in the Online Safety Code within the next legislative cycle.
- Mutual recognition agreements with overseas regulators for enforcement and safety certifications.
- Documented decrease in harmful content accessible in Australia that originates offshore.
- Growth of Australian safety-tech solutions that meet or exceed international standards.

Evidence

- UN Women. (2024). *Normative Advances on Technology-Facilitated Violence Against Women and Girls*, harmonisation and Safety by Design across the digital ecosystem [1].
- Hamburg Declaration. (2023). *Safety by Design Harmonisation Principles* [2].
- Council of Europe. *Lanzarote Convention*, international cooperation on child protection online.
- UK Age Appropriate Design Code and EU Digital Services Act, global benchmarks for child online safety.

References

[1] UN Women. (2024). *Normative Advances on Technology-Facilitated Violence Against Women and Girls*.
[2] Hamburg Declaration. (2023). *Safety by Design Harmonisation Principles*.

7. Other Related Matters: Scams, TFGBV and AI

The current Code underplays scams, technology-facilitated gender-based violence (TFGBV), and AI-driven harms,issues that are rapidly escalating. Scams now exploit the same behavioural, identity and recommendation systems that children use daily. GASA's *State of Scams* report found 63% of adults in Southeast Asia were scammed in the last 12 months, and nearly one in four lost money [1]. These are the very adults supervising children online, often with limited digital literacy and lower confidence in privacy settings.

Technology-facilitated gender-based violence (TFGBV), including grooming, sexual extortion, deepfake abuse, nudification and image-based abuse, increasingly targets minors and women, especially in regional or low-literacy communities. The UN Normative Advances on TF VAWG recognise deepfakes and non-consensual imagery as gender-based violence and call for survivor-led, human-rights due diligence in platforms' safety design [2].

Generative AI amplifies these threats. AI tools can create realistic CSAM or "nudified" images of minors, automate harassment, and generate hyper-targeted scam content. Without strong, mandatory mitigation and oversight, these harms will scale faster than regulators can respond.

Key gaps

- Scams treated as adult-only risk, ignoring how scammers exploit children's accounts or carers' details.
- TFGBV absent from the Code's annexes, no specific obligations for deepfake detection or nudification downranking.

- No AI red-teaming requirement, platforms self-report on AI safety without independent testing.
- Lack of survivor-led training, moderators lack contextual understanding of TFGBV and grooming tactics.

Recommendations

A. Add a TFGBV and Scam Safety Annex to the Code

- Include obligations for deepfake detection, nudification downranking and scam friction measures across search, messaging, and recommendation systems.
- Mandate proactive detection of known scam scripts, phishing domains and impersonation campaigns targeting minors and carers.

B. Fund Survivor-Led Training and Child-Rights Impact Assessments

- Require that moderators and trust-and-safety teams receive survivor-informed training on grooming, sextortion, and TFGBV.
- Make child-rights impact assessments mandatory for all AI features likely to be accessed by minors.

C. Mandate Red-Team Testing of Search and Recommender Systems for Grooming, Sextortion and Fraud

- Independent red-team testing must include AI-driven harm simulations to identify vulnerabilities before deployment.
- Require disclosure of test results and mitigation plans to the regulator.

D. Build Public Awareness and Regional Capacity

- Establish regional scam awareness campaigns integrated with digital literacy programs for carers and older Australians.
- Partner with local councils, NGOs and libraries to provide in-person scam and TFGBV education.

Measures of Success

- Year-on-year reduction in reported TFGBV incidents and scam losses in under-18 cohorts.
- Documented decrease in the time to detect and remove deepfake or nudified content.
- Increased uptake of scam-friction tools and reporting pathways by families and carers.
- Public dashboards showing red-team testing outcomes and AI feature risk ratings.

Evidence

- Global Anti-Scam Alliance. (2023). *State of Scams Southeast Asia*, 63% of adults scammed; nearly one in four lost money [1].
- UN Women. (2024). *Normative Advances on Technology-Facilitated Violence Against Women and Girls*, recognises deepfakes and non-consensual imagery as gender-based violence [2].

- eSafety Commissioner. (2025). *BOSE Periodic Notices*, 1,325% increase in generative-AI CSAM, no provider fully proactive on livestreaming CSEA [3].
- WeProtect Global Alliance. (2023). *Global Threat Assessment*, 360% rise in self-generated sexual images of 7–10-year-olds [4].

References

[1] Global Anti-Scam Alliance. (2023). *State of Scams Southeast Asia*.
[2] UN Women. (2024). *Normative Advances on Technology-Facilitated Violence Against Women and Girls*.
[3] eSafety Commissioner. (2025). *Basic Online Safety Expectations (BOSE) Periodic Notice – CSEA & Sexual Extortion*.
[4] WeProtect Global Alliance. (2023). *Global Threat Assessment*.

References

- eSafety Commissioner. (2025). Basic Online Safety Expectations (BOSE) Periodic Notice – CSEA & Sexual Extortion
- BOSE-full-report-CSEA-sexual-ex…
- Lu, M. et al. (2024). A Content Analysis of Metrics on Online Child Sexual Exploitation and Abuse Used by Online Content-Sharing Services
- 1-s2.0-S0145213424004368-main
- Engelmann, L. et al. (2025). Developing Quality Standards for Community-Based Online Child Sexual Exploitation and Abuse Interventions
- 1-s2.0-S0145213425001991-main
- UN Human Rights Council. (2025). Report of the Special Rapporteur on the Sale and Sexual Exploitation of Children, A/HRC/58/52/Add.1
- A_HRC_58_52_Add.1-EN (2) (1)
- UN Women. (2024). Normative Advances on Technology-Facilitated Violence Against Women and Girls
  normative-advances-on-technolog…
- WeProtect Global Alliance. (2023). Global Threat Assessment
- Global-Threat-Assessment-2023-E…
- Global Anti-Scam Alliance. (2023). State of Scams Southeast Asia
- GASA State of Scam Southeast As…
- Hamburg Declaration. (2023). Safety by Design Harmonisation Principles.

Recommendations Summary

1. Privacy and Data Protection

- Require a Child Rights and Privacy Impact Assessment (CRPIA) for every age-assurance deployment, independently reviewed and published.
- Codify data minimisation, non-reidentification and strict deletion timelines modelled on GDPR and Lanzarote principles.
- Ban secondary use of verification data for targeting, profiling or monetisation.
- Create a publicly searchable Register of Age-Assurance Providers with annual plain-English audit summaries.
- Prefer tokenised age assertions and zero-knowledge proofs so platforms never directly hold sensitive ID data.
- Fund assisted and offline pathways in libraries, councils and NGOs for low-literacy and 55+ users.
- Mandate usability and accessibility testing with low-literacy and 55+ cohorts and publish results.

2. Data Collection and Profiling

- Require privacy-by-default for all services used by minors, including automatic opt-outs from tracking and profiling.
- Extend privacy protections to carers and older Australians who manage online accounts on behalf of children.
- Prohibit selling or sharing age-assurance data with data brokers, marketers or unrelated government programs.

3. Safety by Design and Filtering

- Make Safety by Design mandatory: SafeSearch defaults, blurred previews, TFGBV filters and crisis-card interventions even for non-logged-in traffic.
- Extend filtering and risk controls to autocomplete, recommendation engines and link previews.

- Mandate independent red-team testing of age-assurance and filtering systems, including AI-based harm simulations.
- Require AI-assisted grooming and sextortion detection models with strict privacy safeguards.
- Fund regional pilots to test filters under low bandwidth and low digital literacy conditions.

4. Digital Literacy and Community Support

- Establish a National Digital Literacy & Age-Assurance Support Program for libraries, councils, schools and NGOs.
- Develop plain-language, multilingual curricula covering age-assurance, scam recognition, privacy settings and reporting harmful content.
- Offer micro-credentials and train-the-trainer programs to build local capacity.
- Link grant funding and procurement to demonstrable Safety by Design outcomes.

5. Open Standards and Innovation

- Mandate open parental-control and safety APIs across devices and platforms.
- Create an Australian Safety-Tech Standards Body to govern interoperability.
- Fund a Youth Safety Innovation Lab to co-design safer UX and test emerging tools.

6. Oversight and Transparency

- Establish a Standing Youth Advisory Council under the Online Safety Act.
- Require annual public dashboards reporting on harm reduction, help-seeking and literacy gains.
- Tie industry compliance to measurable improvements, not only box-ticking.
- Commission independent academic and NGO-led evaluations of the Codes every two years.

7. Global Standards Alignment

- Align code provisions with UN CRC General Comment No. 25, the UK Age Appropriate Design Code and the EU Digital Services Act.
- Integrate the 4Cs risk model (Contact, Content, Conduct, Contract) across all code guidance and audits.
- Ratify and implement the Council of Europe Lanzarote Convention on child protection online.
- Establish formal international enforcement partnerships and youth-led peer reviews.
- Incentivise platforms that voluntarily exceed international standards (algorithmic impact assessments, open-sourced safety tools).

8. Scams, TFGBV and AI-Driven Harms

- Add a TFGBV and Scam Safety Annex to the Code, including deepfake detection, nudification downranking and scam friction measures.
- Fund survivor-led training for moderators and make child-rights impact assessments mandatory for all AI features likely to be accessed by minors.
- Mandate red-team testing of search and recommender systems for grooming, sextortion and fraud scenarios.
- Establish regional scam awareness campaigns integrated with digital literacy programs.

In Closing

I strongly support the eSafety Commissioner's leadership and the Statement of Commitment to Children's Rights. Australia has been a pioneer in embedding child safety online and the Commissioner's office has set a global benchmark for evidence-based regulation. This inquiry is an opportunity to strengthen that leadership and ensure that our regulatory framework delivers on its promise of real-world safety for every child, everywhere in Australia.

However, it is equally important to acknowledge the critical gaps and risks that remain. Age-assurance and filtering technologies, if implemented without rigorous privacy safeguards and usability standards, could harm the very people they intend to protect. Regional families, older Australians, carers with low digital literacy, and children in marginalised communities may be left behind or further exposed. Logged-out browsing, "lawful but awful" content, and AI-driven harms, including deepfakes, nudification and scams, demand bold, proactive solutions beyond age restrictions alone.

The Safety by Design approach must become the basic standard rather than an aspiration. This means building privacy, accessibility, and child-rights impact assessments into every layer of the digital ecosystem; mandating independent testing and transparent dashboards; funding community-based digital literacy; and closing the loopholes that allow harmful content to thrive outside traditional moderation pathways.

With my lived experience of being groomed online and my track record of advocacy at the UN, Geneva Peace Week, NCWV, Women 4 STEM, and across regional Australia, I have seen both the devastation of harm and the power of prevention. This submission reflects a prevention-of-harm approach that supports the eSafety Commissioner's work while also urging ambitious, measurable reforms to ensure that regulation delivers not only compliance but true safety outcomes for children and young people.

Australia now has the chance to lead globally by example. By embedding child rights and Safety by Design into every aspect of our online safety framework, we can create a digital environment where children and communities are not just protected but empowered, and where no child is left behind because of their postcode, age, or digital skills.

Thank you for the opportunity to provide this submission to the Senate Environment and Communications Committee. I appreciate the Committee's focus on strengthening

the Online Safety Code and the under-16 social media ban, and I commend the eSafety Commissioner's ongoing leadership in protecting children and young people online.

It is my hope that the evidence and recommendations outlined here will support the Committee's work to ensure that safety by design, privacy, and equity become the baseline for all digital services in Australia. I am grateful for the chance to contribute my lived experience and the insights from Away from Keyboard (AFK) Inc. to help shape a safer, fairer online environment for every child and community.

Sarah Barnbrook

Sarah Barnbrook
Away from Keyboard Inc.
Founder & CEO

Submitted by: Sarah Barnbrook – Founder, Away from Keyboard (AFK) Inc.
Contact: info@afk.org.au +61427203595