alannah & madeline foundation

Online Safety (Basic Online Safety Expectations) Amendment Determination

A submission by the Alannah & Madeline Foundation

February 2024

Contents

Executive summary

The Alannah & Madeline Foundation (the Foundation) welcomes the opportunity to comment on proposed amendments to the Basic Online Safety Expectations (BOSE) Determination 2022. The Determination sets out the Government's minimum expectations of industry with regard to user safety under the Online Safety Act.* Our submission focuses on the items most relevant to our advocacy for the rights of the child.

We welcome the proposal to insert an additional expectation into the BOSE Determination that service providers will take reasonable steps to ensure 'the best interests of the child' are a primary consideration in the design and operation of any service used by or accessible to children. This aligns with the United Nations Convention on the Rights of the Child (UNCRC) General Comment 25 and international policy directions. This expectation could be a 'game-changer' for Australian children online. But it must be backed by clear, child rights-based definitions, child rights impact assessments, expert regulatory guidance, transparent reporting by industry, and alignment with other regulatory developments.

The consultation process also proposes to amend the Determination to specify that any age assurance mechanisms to stop children from accessing Class 2 materials† should be 'appropriate', and to encourage continual improvement by industry in that area. While these proposals are broadly positive, we urge that 'appropriate' approaches to age assurance should reflect classification levels and child rights principles. At present, the consultation paper's framing of the matter has ambiguities which providers might exploit. It is also vital to engage with leaders of the coming Children's Online Privacy Code, and to learn from difficulties faced by creators of the UK Children's Code in relation to age assurance for adult-only sites.

The Foundation reiterates our support for complementary steps: a funded pilot of age assurance technologies to keep pornography away from children; a legislative and regulatory scheme for accreditation and oversight of age assurance providers; and a move away from industry-led codes under the Online Safety Act and towards development of standards by eSafety. While such steps are beyond the scope of the BOSE Determination, they would transform the context in which the Determination functions, lifting its chances of success.

Meanwhile, the consultation process proposes adding several other new items to the Determination which we found especially significant:

- additional expectations that service providers will consider and address end-user safety in relation to generative artificial intelligence (AI) and content recommender systems
- reasonable steps urging service providers to fund staff, systems, tools and processes to support end-user safety
- an additional expectation that service providers will provide controls to help individuals make safer choices (eg. blocking, muting).

While we are broadly supportive of these measures, we believe they could be strengthened by a stronger recognition of the rights of the child.

---

* The BOSE Determination applies to social media services, relevant electronic services, and designated internet services. A social media service is a service that has the sole or primary purpose of enabling online social interactions between end-users, where end-users can also link to other end-users and post material on the service. A relevant electronic service is a service that allows end-users to communicate with other end users by means of email, instant messaging, SMS, MMS, chat services or online games. A designated internet service is a service that allows end-users to access material on the internet using an internet carriage service or that delivers material to persons by means of an internet carriage service (e.g. websites).
† Class 2 material typically refers to material which has been classified, or would likely be classified, as X18+ or R18+ by the Classification Board under the Classification (Publications, Films and Computer Games) Act 1995. For full definition, see the Online Safety Act, section 107.

This could mean:

- explicit recognition of the need to consider and address threats to the rights of children in relation to generative AI capabilities, given children's vulnerabilities, unique experiences of AI, and risk of neglect by AI policy-makers
- particular emphasis on the need for industry to invest in measures to counter child sexual exploitation and abuse, given the prevalence, growth and severity of the issue
- reiteration that privacy and safety settings should be set to the highest level by default for children – a pressing concern given that the first round of industry codes and standards under the Online Safety Act (Class 1A and 1B material) appeared to undermine this approach
- widening the definition of 'recommender systems' to include systems that recommend friends or followers, not just content, thus broadening the scope of protections for children.

The BOSE Determination functions within the scope of the Online Safety Act, which focuses strongly on 'content' risks. As such, the Determination may not be well-placed to address other risks to children in the digital environment eg. risks of contact, conduct, contract, and compulsion. However, the Determination still has a crucial part to play. We hope to see it strengthened in ways which will make a meaningful difference to building a better digital environment for children.

About us

The Foundation was established the year after the Port Arthur tragedy, by Walter Mikac AM in memory of his two young daughters, Alannah and Madeline. Our vision is that all children and young people are safe, inspired and have freedom to flourish.

Over the last 25 years our work has grown and evolved but our purpose remains the same. We have three program streams:

- Safe and Strong: recovering and healing from trauma. Linked to our origin story, we have a specialist trauma recovery and therapy service for children who have experienced significant trauma. This has grown in recent years to include working with early childcare providers, kindergartens, and now primary schools to help them build their trauma informed capability and practices. Most of our work in trauma healing and recovery is Victorian based, with our therapists and consultants working from our client's homes and places of work.
- Safe and Strong: building positive digital citizens. The Foundation supports schools, educators, families and communities nationally to build digital skills and competencies to develop a generation of safe and strong digital citizens. For over 12 years the Foundation has delivered eSmart, an initiative designed to empower children (3 - 18 years) to be safe and responsible online. It encompasses a range of learning tools and resources to help students build essential digital and media literacy skills, so they can thrive online.
- Safe and Strong: bringing children's rights to life. As a rights-based organisation, this is our policy and advocacy work. Since inception, we have advocated for firearms safety, and we convene the Australian Gun Safety Alliance. In other key policy matters related to our programs, we work closely with the Office of the eSafety Commissioner, the Prime Minister's National Office for Child Safety and other major agencies such as the Australian Federal Police.

In 2018, we partnered with Kate and Tick Everett, after the tragic suicide of their daughter, Dolly. With them we worked to establish Dolly's Dream.

- Safe and Strong: Dolly's Dream, changing the culture of bullying. The purpose is the same, but the programs and services (Parent Hub, telephone help line, school, and community workshops etc.) are specifically designed for remote, rural, and regional families and communities, to meet their unique needs and contexts.

Recommendations

1. **Proposed addition 6(2A):** Adopt the proposed addition 'The provider of the service will take reasonable steps to ensure that the best interests of the child are a primary consideration in the design and operation of any service that is used by, or accessible to, children.' Clarify that a child is defined in line with the Online Safety Act as an individual who has not yet reached 18 years, and that 'the best interests of the child' is defined in line with Article 3 of the United Nations Convention on the Rights of the Child.

2. **Proposed addition 6(2A):** Clarify that the 'best interests of the child' obligation applies to any service likely to be accessed by children. Create guidance to enable assessment of which services are in scope. When services are deemed not to be in scope, there should be documentation and explanation – see the approach taken by the leaders of the UK Children's Code. Services should not be able to self-exclude on grounds that they are not aimed specifically at children or that they contain age-gating mechanisms (which may be weak). Nor should services be able to shift responsibility inappropriately onto public services eg. edtech providers positioning schools as holding primary responsibility for the safety of their digital products.

3. **Proposed addition 6(2A):** Require that services must carry out a child rights impact assessment as part of treating the best interests of the child as a primary consideration.

4. **Proposed addition 6(2A):** Work with leaders of the coming Children's Online Privacy Code – presumably the office of the Australian Privacy Commissioner and the Attorney-General's Department – to ensure alignment in treating the best interests of the child as a primary consideration and determining 'appropriate' approaches to age assurance. Consider, too, the approach taken by the UK Children's Code, flagged as a model for Australia.

5. **Proposed addition 6(2A):** Engage with the National Children's Commissioner to resource expert, appropriate guidance for service providers in assessing and upholding the 'best interests of the child' as a primary consideration.

6. **Reasonable step 6(3)(b):** Clarify how the BOSE Determination's reasonable step of high default privacy and safety settings for services used by children under 18 aligns with the position of the recent industry code for social media services and the draft industry standard for relevant electronic services (Class 1A and 1B materials). The code and draft standard limit the requirement for high default privacy settings to under-16s instead, which we consider inadequate.

7. **Proposed additions 6(3)(g) and (h):** Adopt the proposed additions encouraging service providers to invest in 'staff, systems, tools and processes to action reports and complaints within a reasonable time' and 'systems, tools and processes to improve the prevention and detection of material or activity on the service that is unlawful or harmful'. Add a further item noting the urgency of investing in measures to prevent and address child sexual exploitation and abuse (CSEA), including via generative AI. Ideally, service providers would aim for the 'maturity' measure of innovative solution development by industry articulated in the WeProtect Model National Response to preventing and tackling online CSEA.

8. **Proposed addition 8A:** Adopt the proposed addition 'provider will take reasonable steps regarding generative artificial intelligence capabilities.' Add a further item requiring services which use or enable

the use of generative AI to treat the best interests of the child as a primary consideration as regards any generative AI capabilities likely to have an impact upon children.

9. **Proposed addition 8A:** Add a reasonable step urging services to put in place clear, accessible mechanisms for individuals to make complaints or enquiries about generative AI capabilities functioning to create, promote or distribute material, or enable activity, that is unlawful or harmful.

10. **Proposed addition 8B:** Adopt the proposed addition 'provider will take reasonable steps regarding recommender systems'. Clarify that this measure is intended to address not only recommender systems which suggest content to end-users, but also those which suggest contacts.

11. **Proposed amendment 12(2)(a):** Clarify that 'appropriate' age assurance mechanisms to prevent children from accessing Class 2 material should be effective, lawful (eg. reflecting the adult-only classifications of X18+ and R18+) and proportionate to the likelihood of children encountering Class 2 material on a particular service. Mechanisms should uphold the principle of data minimisation and the rights of the child, including children's rights to freedom from arbitrary or unlawful interference with their privacy and to protection from all forms of sexual exploitation. A child rights approach would be further advanced by meaningful engagement with children and young people to inform the design and implementation of age assurance mechanisms, for example, as part of a funded pilot as recommended by eSafety.

12. **Proposed amendment 12(2)(a):** Clarify that tokenistic approaches to age-assurance, such as allowing children to self-declare their age, should not be viewed as 'reasonable steps' for online services which are age-restricted by law and/or not intended or appropriate for children to use in any case eg. pornography sites.

13. **Proposed addition 18A:** Adopt the proposed addition 'provider will publish transparency reports'. Require that key terms be defined clearly, accurately and consistently, including 'effectiveness', 'harms' and 'responsiveness'. Ideally, meaningful metrics would be agreed upon with eSafety.

14. **Proposed addition 18A:** Add a further item requiring relevant service providers to publish in their transparency reports information about their child rights impact assessments and how they have assessed, determined and acted on the best interests of the child as a primary consideration.

The best interests of the child as a primary consideration

We welcome the proposal to create a new additional expectation 6(2A): 'The provider of the service will take reasonable steps to ensure that the best interests of the child are a primary consideration in the design and operation of any service that is used by, or accessible to, children.'

This would be an important step forward, with potential to significantly improve children's experiences online.

To deliver the best outcomes for children, we call for five supplementary steps: definition of key terms; expert guidance; child rights impact assessment; transparency reporting; and alignment with other reforms.

Clear, agreed terminology is important. We call for clarification that:

- A 'child' is defined in line with the Online Safety Act and the United Nations Convention on the Rights of the Child (UNCRC) as an individual who has not reached 18 years.
- 'The best interests of the child' aligns with Article 3 of the UNCRC.
- '[A]ny service that is used by, or accessible to, children' means any service likely to be accessed by children. For example, providers may not self-exclude on grounds that their service is not targeted at children or that it is 'protected' by nominal age-gating (which may be low-quality).

Without clear definitions, service providers might interpret their obligations in ways that lead to weaker protection of children's rights. For example, the recent creation of industry codes and standards for Class 1A and 1B material saw some safety measures mandated only for 'young children' under 16, thus lowering the bar of protection. Meanwhile, the community holds many different understandings of 'the best interests of the child' – a complex term to interpret, even for experts. It is also worth noting the extensive work undertaken in the UK to clarify which online services were in scope of their Children's Code (ie. most of them) – hinting that many services would have opted out unless their obligations were made unambiguous.

It seems likely to us that service providers will need expert guidance in assessing and upholding the best interests of the child.[1] The concept of 'the best interests of the child' aims to ensure the full and effective enjoyment by children of all their rights, as well as the holistic development of the child. According to General Comment 25 of the UNCRC ('On children's rights in relation to the digital environment'), states should 'have regard for all children's rights, including their rights to seek, receive and impart information, to be protected from harm and to have their views given due weight'.[2] General Comment 14 of the UNCRC ('On the right of the child to have his or her best interests taken as a primary consideration') states that the following elements are to be taken into account: the child's views; the child's identity; preservation of the family environment and maintaining relations; care, protection and safety of the child; the child's situation of vulnerability; the child's right to health; and the child's right to education.[3]

Treating the best interests of the child as 'a primary consideration' means decision-makers recognise children's particular vulnerabilities and the need to give high priority to children's best interests, not treating them as just one of several considerations.[4]

To create appropriate guidance, we encourage engaging the National Children's Commissioner, in line with the advice of UNCRC General Comment 25 that states should involve national and local bodies that oversee the fulfilment of children's rights in ensuring that all actions regarding the provision, regulation, design, management and use of the digital environment treat the best interests of the child as a primary consideration.[5]

Expert guidance may also be needed to clarify which services are required to treat the best interests of the child as a primary consideration. By way of comparison, the creators of the UK Children's Code (which also has 'the best interests of the child' as a central premise) provided guidance to help services estimate whether children's access was likely. Services were encouraged to consider things like research findings about children's access; behavioural data implying use by children; the rigour of any age verification mechanisms; and content or advertising on the service aimed at or appealing to children. Services which decided they were not likely to be accessed by children were instructed to document their reasons.[6] The Code creators also released updated guidance confirming that private edtech providers can have responsibilities under the Code, after concerns were raised about providers avoiding this (despite profiting from children's data) by contractually positioning schools as the responsible 'data controllers'.[7] It is important for Australia to avoid any similar pitfalls.

To treat 'the best interests of the child' as a primary consideration, service providers will need to undertake some process of assessment and actioning. UNCRC General Comment 25 specifies: 'States parties should require the business sector to undertake child rights due diligence, in particular to carry out child rights impact assessments and disclose them to the public, with special consideration given to the differentiated and, at times, severe impacts of the digital environment on children.'[8] (Our emphasis.)

No such assessments are flagged in the BOSE Determination. Some services may engage in related forms of risk assessment elsewhere. For example, the BOSE Determination 12(2)(b) lists child safety risk assessments as a 'reasonable step' in relation to protecting children from Class 2 material. Meanwhile, eSafety's Safety By Design model expects participating providers to document their risk management and impact assessments.[9] And the first set of industry codes requires social media services to assess the risk that Class 1A and 1B material will be accessed, distributed, or stored on their service; notify eSafety of their risk profile and the reasons behind it; and be able to demonstrate that their risk assessment is based on reasonable criteria.[10]

However, these risk assessment processes are either voluntary, limited in scope, or both. Without a dedicated child rights impact assessment, we contend it is difficult for providers to state credibly that they are upholding the best interests of the child.

It is important the community has opportunities to understand the decisions made by service providers regarding 'the best interests of the child'. To this end, we believe the proposed amendment 18A to the BOSE Determination ('Additional expectation – provider will publish transparency reports') should include an expectation that service providers will, as relevant, publish information about their child rights impact assessment and how they have determined and acted on the best interests of the child.

Finally, we note the importance of alignment between the amended BOSE Determination and the promised Children's Online Privacy Code. The Determination addresses 'the best interests of the child' within the parameters of the Online Safety Act, which focuses largely on 'unlawful or harmful' content.‡ It is likely the Children's Online Privacy Code will have a similar scope to the UK Children's Code,[11] which focuses on 'the best interests of the child' in relation to the way online service providers handle children's personal information – a different but overlapping issue. For example, the UK Children's Code requires services to switch off geolocation options by default, provide an obvious sign for children when location tracking is active, and ensure functions which make a child's location visible to others will default to 'off' at the end of the session.[12] Safety and privacy are related; alignment will be key.

Note: it may also be necessary for leaders of the BOSE Determination to clarify that upholding the best interests of the child does not necessarily equate with implementing age assurance mechanisms. In community discussions about this consultation process, we have noticed conflation and confusion between the proposed additions 6(2A) [the best interests of the child] and 12(2)(a) and (b) [age assurance].

Age assurance

A core expectation of the BOSE Determination is that digital service providers will take reasonable steps to ensure technological or other measures are in effect to prevent access by children to Class 2 material provided on the service.

This consultation process proposes two amendments. Firstly, it proposes that the reasonable step of implementing age assurance mechanisms should now read 'implementing appropriate age assurance mechanisms'. Another reasonable step is proposed for industry: 'continually seeking to develop, support or source, and implement improved technologies and processes for preventing access by children to Class 2 material'.

Overall, we support these additions. However, we would welcome more detail about what is meant by 'appropriate' mechanisms and how these approaches will align with wider regulatory changes.

In relation to the digital environment in general, we see value in making approaches to age assurance proportionate to the likelihood of children experiencing risk and harm in different online spaces. However, we believe greater clarity is needed about the approach to age assurance in relation to Class 2 material proposed for the updated BOSE Determination. The consultation paper (p.11) states:

> 'The inclusion of the word "appropriate" signals that age assurance mechanisms to prevent children's access to Class 2 material should be calibrated to the level of risk and harm of the material. This means that in some instances, asking users to self-report their age or date of birth may provide an effective signal or barrier to unintentional access by children, while in other instances, services will be expected to establish a user's age with a greater level of certainty that is appropriate for the level of risk of the material they may access on the service.'

‡ Cyber bullying material targeted at children; cyber abuse material targeted at adults; material depicting abhorrent violent conduct; seriously illegal content; restricted content; and intimate imagery used without consent.

We are concerned that this might be interpreted as implying that some Class 2 materials pose less risk to children and that self-declaration of age or date of birth is a sufficient child safety measure for some digital spaces where Class 2 material is present.

Surely 'the level of risk and harm of the material' has already been established through the classification of Class 2 material as X18+ or R18+ ie. legally restricted to adults only.[13] This position is backed by high levels of community concern about children's exposure to pornography and the support of three-quarters of Australian adults for some form of age assurance to keep pornography away from children.[14] Self-reporting age or date of birth seems an inadequate safety measure where Class 2 material is present, not something that should be promoted as a reasonable step.[§]

Possibly the consultation paper's intended meaning was different. It may have meant that 'light-touch' approaches to age assurance can be appropriate for online spaces where the likelihood of children encountering Class 2 material is low. If this was the intent, we call for clarification that 'risk' in this context refers to a child's risk of finding Class 2 material on a site, not the riskiness of the material itself.

Otherwise, we would echo the position of the UK Information Commissioner's Office: when services are age-restricted by law, or when they are not intended or appropriate for children to use in any case, the focus should be on preventing access by children, while still preserving children's other rights eg. to privacy.[15]

Again, it is important to keep in mind the wider regulatory context. We supported eSafety's recommendation that age assurance technologies be trialled in Australia based on lessons from existing pilots, before being mandated.[16] This would help shift a longstanding problem of tokenistic or ineffective barriers to age-restricted content, goods and services online.[17] We were disappointed that the Australian Government postponed a decision on this until after the development of Class 2 industry codes[18] and hope to see progress in 2024.

As we have stated before, the Foundation does not support the industry-led approach to code creation under the Online Safety Act. It would be difficult for the community to have confidence in a code intended to keep Class 2 material away from children if the code were drafted by stakeholders with a strong commercial interest in growing their young user base. We would prefer that standard development was led by eSafety.

Presumably the topic of age assurance will also be addressed by the Children's Online Privacy Code. The code is envisaged as being similar in scope of the UK Children's Code,[19] which directs service providers to either establish end-user age with a level of certainty appropriate to the risks and freedoms of children that arise from the service's processing of their data (aligning with the UK GDPR principles for data protection), or else apply the Code's standards to all users.[20] In 2022, the Information Commissioner's Office (ICO) clarified that while adult-only sites should not be accessible to children, they may be within scope of the Code if they are likely to be accessed by children. This decision, accompanied by guidance for industry, was made in recognition of the fact that many children do access adult-only sites, with significant risks attached.[21]

As eSafety has noted, Australia is also in need of a suitable legislative and regulatory scheme for the accreditation and oversight of age assurance providers.[22] The development of such a scheme would address a recommendation of the 2020 parliamentary report 'Protecting the age of innocence', which called for development of standards for online age verification for age-restricted products and services.[23] The then-government agreed to this in principle.[24]

Age assurance should not mean that children lose their right not to be subjected to arbitrary or unlawful interference with their privacy. It is our understanding that related work has been undertaken to inform Australia's Digital ID Bill. Support has been growing for an approach to online verification that is privacy-respecting, secure, convenient and reusable, able to verify relevant attributes (eg. age) without passing on

identifying personal details to a commercial platform. A key principle is strict separation between the provider that needs confirmation of an attribute (eg. that an individual is over 18) and the provider that conducts the

[§] In their Roadmap for Age Verification, eSafety stated 'age gates based on self-declaration present no barrier to those who want to evade them'. p.44

verification. As such approaches take shape, they will provide new possibilities for approaching age assurance effectively and ethically.

Generative artificial intelligence

While generative AI offers many potential benefits to children, it can be developed and deployed in ways which violate children's rights. We welcome the proposal to insert a new additional expectation, 8A, that providers of services which use or enable the use of generative AI capabilities will take reasonable steps to consider end-user safety; incorporate safety measures into the design, implementation and maintenance of AI capabilities; and proactively minimise the extent to which generative AI capabilities may be used to produce unlawful or harmful material.

At present, the proposed addition 8A does not mention reporting or complaint mechanisms for individuals. Providing accessible, effective ways for individuals to raise concerns is an important part of addressing user safety and speaks to the right of children to have a voice in matters affecting them. A new reasonable step might be added to this effect.

Moreover, at present the proposed additional expectation about AI does not mention children. We believe the best interests of the child should be recognised explicitly here, for three reasons:

- Children's innate vulnerability due to their early stage of life and development.
- The unique experiences of today's children, who are affected by AI from a very early age in ways which adults do not always anticipate or understand.
- The 'immature' approach to policy-making on the topic of AI and children identified in a 2020 UNICEF review of 20 national AI strategies. Most AI strategies made only cursory mention of children, usually in the context of education and future employment, with inadequate attention paid to the ramifications for children's other rights.[25]

Perhaps the most egregious threats posed by generative AI relate to child sexual exploitation and abuse. Generative AI can be used by offenders to return realistic imagery; script sextortion or grooming interactions; suggest methods for abusing a child or finding abuse material; mask child sexual abuse material to avoid detection; and pool information on how to destroy evidence and evade detection. AI-generated imagery also creates many difficulties for law enforcement. Consequently, WeProtect Global Alliance concludes:

> 'Generative AI represents a paradigm shift that underlines the need for Safety by Design … There is currently no evidence that child safety has been integrated into the design and rollout of generative AI services. In the context of the sustained increase in reported child sexual exploitation and abuse online, E2EE [end-to-end encryption] adoption and emergent technologies such as AI signal a critical juncture at which urgent, widespread implementation of Safety by Design represents the only viable route to turning the tide on current trends.'[26]

We recognise that the proposed additional expectation 6(2A) already specifies that the best interests of the child should be a primary consideration in the design and operation of any service that is used by, or accessible to, children. However, we feel the section on generative AI should also include the 'best interests of the child' as a primary consideration. Threats posed to children by generative AI do not always happen in digital environments accessed voluntarily by children, and these threats may derive from the behaviour of adults or the technology itself. Therefore 6(2A) may not go far enough. (We suspect services will likely address their obligations under 6(2A) by focusing on factors relevant to children's own direct interactions with their platforms.) Setting 'the best interests of the child' as a primary consideration could also help to encourage investment in generative AI that can help to detect, deter and remove CSEA.

Resourcing for online safety measures

We welcome the proposal to insert two reasonable steps, 6(3)(g) and (h) into the BOSE Determination, to encourage digital service providers to have 'staff, systems, tools and processes to action reports and complaints within a reasonable time' and 'invest in systems, tools and processes to improve the prevention and detection of material or activity on the service that is unlawful or harmful'.

The consultation paper states that this would address a gap in the current Determination, namely: 'it does not expressly encourage investment of resources for the purpose of achieving better safety outcomes for end-users'. (p.12) We concur: without adequate resourcing, regulation becomes meaningless.

Ideally, we would like to see specific recognition of the need to invest in measures to prevent and address child sexual exploitation and abuse (CSEA), which eSafety has recognised as 'a particularly high risk and high harm issue that has seen sustained growth'.[27]

In 2022 and 2023, eSafety's nonperiodic reporting notices found that online services' use of measures to prevent and address CSEA was very uneven. There was 'no common baseline' for safety protections between different services, even those owned by the same parent company. Not all companies used available technologies that would have enabled them to detect and remove images containing CSEA; detect CSEA in video calls and livestreams; identify URLs which linked to known CSEA material; and detect language indicative of likely CSEA activity, such as sexual extortion. In user-governed online communities, enforcement of rules was inconsistent, with risks attached to the use of volunteer moderators and administrators. The length of time taken to respond to reports of child sexual exploitation and abuse varied significantly between services. Meaningful age assurance on supposedly age-limited platforms was rare.[28]

In light of this, we would like to see relevant service providers urged to invest – proactively, meaningfully, and on an ongoing basis – in solutions that strengthen their ability to detect, deter and disrupt CSEA, including 'first generation' CSEA.

On a related matter, the need for resource investment is not stated explicitly in the proposed new expectation 8A, which addresses generative AI capabilities. Given the threats posed by the evolution of generative AI, a more explicit link between the proposed amendments 6(3)(g)/(h) and 8A would be welcome.

Ideally, we would like to see service providers align with the 'maturity' measure of innovative solution development by industry set out in the Model National Responses (WeProtect Global Alliance and UNICEF):

> 'Innovative technological solutions that demonstrably enhance existing approach to preventing and tackling child sexual exploitation and abuse are consistently and effectively developed, scaled up, monitored and updated. Industry actively finances and prioritises tech solutions that are compatible with children's rights and safety online.'[29]

End-user controls

A new additional expectation is proposed for the BOSE Determination, 6(5): that service providers will take reasonable steps to provide end-users with controls for choice and autonomy. Reasonable steps include controls to enable end-users to block, mute, opt in or out of content, and change privacy and safety settings.

This would be a positive step. However, as we have stated elsewhere, we maintain the settings of end-users aged under 18 should be set to the highest levels of privacy and safety by default.

Such an approach would be in line with community values: 92% of Australian parents support default privacy settings for children being set to high.[30] But parents' own levels of digital literacy are variable and many struggle to help their children make safe choices in an unsafe online world.[31] Requiring high levels of privacy and safety by default would shift the focus away from individual families and towards industry's own obligations. Codes to regulate the handling of children's data in the UK, Sweden, the Netherlands, Ireland

and France all highlight the value of high privacy settings by default.[32] Options presented as default are more likely to be chosen, either actively due to consumer bias or passively due to consumers making no choice.[33] The UK Information Commissioner concluded 'Many children just accept whatever privacy settings they are given and never change them.'[34]

The BOSE Determination seems to concur, stating in 6(3)(b) that a reasonable step for service providers is to ensure that 'if a service or a component of a service (such as an online app or game) is targeted at, or being used by, children (the children's service) – ensuring that the default privacy and safety settings of the children's service are robust and set to the most restrictive level'.

It is concerning, therefore, that the industry code for social media services and the draft industry standard for relevant electronic services under the Online Safety Act (class 1A and 1B materials) require high default privacy settings only for under-16s, not under-18s.[35] This means a lower threshold of protection for Australian children than that available in several comparable countries.[36]

The code and standard define under-16s as 'a young Australian child' – a term not present in the Online Safety Act or the BOSE Determination. Thus, we would query how well the code and standard uphold the relevant legislation and guidance.

Recommender systems

The consultation process proposes to insert an additional expectation into the BOSE Determination, 8B: 'provider will take reasonable steps regarding recommender systems'. This measure would require providers which use recommender systems to take reasonable steps to consider end-user safety, incorporate safety measures, and minimise 'amplification of material or activity on the service that is unlawful or harmful'.

While this proposal is promising, it appears to limit the definition of recommender systems to those which 'prioritise content or make personalised content suggestions to users' (consultation paper, p.8).

But content is not the only focus of automated systems which make suggestions to individuals based on their previous activities and the behaviour of purportedly similar accounts. Recommender systems also suggest connections with new friends or followers. Risks also exist in relation to privacy and consumer rights eg. data handling, targeted advertising, prioritising paid-for content.[37]

Thus, recommender systems pose risks related not only to content, but also to contact and contract – as well as compulsion, if we recognise that their purpose is to maximise engagement.

The BOSE Determination cannot tackle the full range of risks posed by recommender systems, given the parameters of the Online Safety Act, which focuses strongly on content. However, we suggest the risks of friend / follower recommender systems could be addressed viably here. The Determination does include a framing of 'unlawful or harmful material or activity' (our emphasis) and contact between individuals is key to several issues the Online Safety Act was created to address: cyber bullying, image-based abuse, and the spread of Class 1 and 2 material. The eSafety Commissioner has recognised the risks posed to children by recommender systems that make friend/follower suggestions of potentially dangerous adults.[38]

Transparency reporting

We welcome the proposal to insert an additional expectation, 18A, that service providers will publish regular transparency reports with information specific to Australia regarding: '(a) the service's enforcement of its terms of use, policies and procedures and standards of conduct…; (b) the safety tools and processes deployed by the service … and their effectiveness; (c) metrics on the prevalence of harms, reports and complaints, and the service's responsiveness; and (d) the number of active end-users of the service in Australia (including children) each month during the relevant reporting period.'

This is a positive step towards greater accountability. As stated earlier, we believe services should also be required to report on how they have acted to uphold the best interests of the child.

Furthermore, it is important that to have clear, consistent, accurate definitions of the key terms against which services will report eg. 'effectiveness', 'harms,' 'responsiveness'. Ideally, meaningful metrics would be agreed upon with eSafety.

We would welcome the opportunity to discuss any of these matters further. Please contact:

Sarah Davies AM, CEO
[redacted]

Ariana Kurzeme, Director, Policy & Prevention
[redacted]

Dr Jessie Mitchell, Manager, Advocacy
[redacted]

[1] Dr Jonathan Collinson, Jen Persson, 'What Does The "Best Interests of the Child" Mean for Protecting Children's Digital Rights? A narrative literature review in the context of the ICO's Age Appropriate Design Code,' *Communications Law*, 27(3), 2022 https://eprints.whiterose.ac.uk/194985/3/Communications%20Law%20%20BIC%20in%20digital%20rights%20-%20Revisions%2028-2-22.pdf

[2] United Nations Convention on the Rights of the Child (UNCRC), 'General comment No. 25 (2021) on children's rights in relation to the digital environment,' 2021, p.3, https://tbinternet.ohchr.org/_layouts/15/treatybodyexternal/TBSearch.aspx?Lang=en&TreatyID=5&DocTypeID=11

[3] UNCRC, 'General comment No. 14 (2013) on the right of the child to have his or her best interests taken as a primary consideration,' 2013, pp.9, 13-17 https://www2.ohchr.org/english/bodies/crc/docs/gc/crc_c_gc_14_eng.pdf

[4] UNCRC, 'General comment No. 14', p.10

[5] UNCRC, 'General comment No. 25', p.3

[6] Information Commissioner's Office (ICO), 'Introduction to the Children's code', accessed January 2024, https://ico.org.uk/for-organisations/uk-gdpr-guidance-and-resources/childrens-information/childrens-code-guidance-and-resources/introduction-to-the-childrens-code/; ICO, ''Likely to be accessed' by children – FAQs, list of factors and case studies,' accessed January 2024, https://ico.org.uk/for-organisations/uk-gdpr-guidance-and-resources/childrens-information/childrens-code-guidance-and-resources/likely-to-be-accessed-by-children/ ; ICO, 'Services covered by this code,' accessed January 2024, https://ico.org.uk/for-organisations/uk-gdpr-guidance-and-resources/childrens-information/childrens-code-guidance-and-resources/age-appropriate-design-a-code-of-practice-for-online-services/services-covered-by-this-code/

[7] Ayça Atabey, Louise Hooper, Sonia Livingstone and Kruakae Pothong, Digital Futures Commission, 'A step towards clarity: welcoming ICO's new guidance for EdTech on the Age Appropriate Design Code,' July 17, 2023, https://digitalfuturescommission.org.uk/blog/a-step-towards-clarity-welcoming-icos-new-guidance-for-edtech-on-the-age-appropriate-design-code ; Emma Day, Kruakae Pothong, Ayca Atabey, Sonia Livingstone, 'Who controls children's education data? A socio-legal analysis of the UK governance regimes for schools and EdTech', Learning, Media and Technology, December 2022, https://www.tandfonline.com/doi/full/10.1080/17439884.2022.2152838; Digital Futures Commission, 'A Blueprint for Education Data: Realising children's best interests in digitised education,' March 2023, https://digitalfuturescommission.org.uk/wp-content/uploads/2023/03/A-Blueprint-for-Education-Data-FINAL-Online.pdf ; Information Commissioner's Office, 'The Children's code and education technologies (edtech),' 30 May 2023, https://ico.org.uk/for-organisations/uk-gdpr-guidance-and-resources/childrens-information/childrens-code-guidance-and-resources/the-children-s-code-and-education-technologies-edtech/

[8] UNCRC, 'General comment No. 25,' p.7

[9] eSafety, 'Safety By Design: Principles and background,' accessed February 2024, https://www.esafety.gov.au/industry/safety-by-design/principles-and-background
[10] Onlinesafety.org.au, Schedule 1 – Social Media Services Online Safety Code (Class 1A and Class 1B Material), 16 June 2023, pp.5-7 https://onlinesafety.org.au/codes/ ; Onlinesafety.org.au, Consolidated Industry Codes of Practice for the Online Industry (Class 1A and Class 1B Material) Head Terms, 12 September 2023, p.13 https://onlinesafety.org.au/codes/
[11] Australian Government, 'Government Response: Privacy Act Review Report,' 2023, p.30 https://www.ag.gov.au/sites/default/files/2023-09/government-response-privacy-act-review-report.PDF
[12] ICO, 'Geolocation,' accessed January 2024, https://ico.org.uk/for-organisations/uk-gdpr-guidance-and-resources/childrens-information/childrens-code-guidance-and-resources/age-appropriate-design-a-code-of-practice-for-online-services/10-geolocation/
[13] Australian Government, *Online Safety Act 2021*, No.76, 2021, p.93
[14] eSafety, 'Roadmap for age verification and complementary measures to prevent and mitigate harms to children from online pornography,' March 2023, https://www.esafety.gov.au/sites/default/files/2023-08/Roadmap-for-age-verification_2.pdf; Parliament of the Commonwealth of Australia, 'Protecting the age of innocence: Report of the inquiry into age verification for online wagering and online pornography,' House of Representatives Standing Committee on Social Policy and Legal Affairs, February 2020, pp.35-39 https://parlinfo.aph.gov.au/parlInfo/download/committees/reportrep/024436/toc_pdf/Protectingtheageofinnocence.pdf;fileType=application%2Fpdf
[15] ICO, 'UK Information Commissioner's opinion: Age Assurance for the Children's Code', 14 October 2021, https://ico.org.uk/media/about-the-ico/documents/4018659/age-assurance-opinion-202110.pdf ; ICO, 'Services covered by this code'; ICO, ''Likely to be accessed' by children'
[16] eSafety, 'Roadmap for age verification'
[17] Svetlana Smirnova, Sonia Livingstone and Mariya Stoilova, 'Understanding of user needs and problems: A rapid evidence review of age assurance and parental controls in everyday life (D2.4a),' euConsent project, 2021, https://euconsent.eu/download/understanding-of-user-needs-and-problems-a-rapid-evidence-review-of-age-assurance-and-parental-controls/
[18] Australian Government, 'Response to the Roadmap for Age Verification,' August 2023, https://www.infrastructure.gov.au/department/media/publications/australian-government-response-roadmap-age-verification
[19] Australian Government, 'Government Response: Privacy Act Review Report', p.30
[20] ICO, 'Introduction to the Children's Code'; ICO, 'UK Information Commissioner's opinion: Age Assurance for the Children's Code'
[21] ICO, '"Children are better protected online in 2022 than they were in 2021" ', 2 September 2022, https://ico.org.uk/about-the-ico/media-centre/news-and-blogs/2022/09/children-are-better-protected-online-in-2022-than-they-were-in-2021/ ; ICO, 'Children's code evaluation', March 2023, https://ico.org.uk/media/about-the-ico/documents/childrens-code/4025494/childrens-code-evaluation-report.pdf; ICO, 'Likely to be accessed guidance: impact assessment', 2023, https://ico.org.uk/media/4025881/ltba-guidance-impact-assessment.pdf . Also: UK Information Commissioner's Office, ''Likely to be accessed' by children'; ICO, 'Services covered by this code'
[21] ICO, 'Introduction to the Children's Code'
[22] eSafety, 'Roadmap for age verification'
[23] Parliament of the Commonwealth of Australia, 'Protecting the age of innocence'
[24] Parliament of the Commonwealth of Australia, 'Protecting the age of innocence'
[25] UNICEF, 'National AI strategies and children: reviewing the landscape and identifying windows of opportunity,' 2020, https://www.unicef.org/globalinsight/media/1156/file . Also UNICEF, 'Policy guidance on AI for children,' 2.0, November 2021, https://www.unicef.org/globalinsight/media/2356/file/UNICEF-Global-Insight-policy-guidance-AI-children-2.0-2021.pdf
[26] WeProtect Global Alliance, 'Global Threat Assessment 2023: Assessing the scale and scope of child sexual exploitation and abuse online, to transform the response,' 2023, https://www.weprotect.org/wp-content/plugins/pdfjs-viewer-shortcode/pdfjs/web/viewer.php?file=https://www.weprotect.org/wp-content/uploads/Global-Threat-Assessment-2023-English.pdf&attachment_id=384869&dButton=true&pButton=true&oButton=false&sButton=true#zoom=0&pagemode=none&_wpnonce=fea2469cda

[27] eSafety Commissioner, 'Basic Online Safety Expectations: summary of industry responses to mandatory transparency notices,' October 2023, https://www.esafety.gov.au/sites/default/files/2023-10/Full-transparency-report-October-2023.pdf
[28] eSafety Commissioner, 'Basic Online Safety Expectations: summary of industry responses to mandatory transparency notices,' Oct 2023; eSafety Commissioner, 'Basic Online Safety Expectations: summary of industry responses to the first mandatory transparency notices,' December 2022, https://www.esafety.gov.au/sites/default/files/2022-12/BOSE%20transparency%20report%20Dec%202022.pdf
[29] WeProtect Global Alliance and UNICEF, 'The Model National Response Maturity Model,' accessed January 2024, https://www.weprotect.org/model-national-response/
[30] Office of the Australian Information Commissioner, 'Australian Community Attitudes to Privacy Survey,' Canberra, 2023, https://www.oaic.gov.au/__data/assets/pdf_file/0025/74482/OAIC-Australian-Community-Attitudes-to-Privacy-Survey-2023.pdf
[31] Australian Centre to Counter Child Sexual Exploitation, 'Understanding community awareness, perceptions, attitudes and preventative behaviours,' Research report, 2020, https://accce.prod.acquia-sites.com/sites/default/files/2021-02/ACCCE_Research-Report_OCE.pdf
[32] 5Rights Foundation, 'Approaches to children's data protection': a comparative international mapping,' October 2022, https://5rightsfoundation.com/Approaches-to-Childrens-Data-Protection---.pdf
[33] Australian Government, Behavioural Economics Team, 'Harnessing the Power of Defaults,' Governance note, https://behaviouraleconomics.pmc.gov.au/sites/default/files/resources/harnessing-power-defaults.pdf
[34] ICO, 'Protect children's privacy by default', February 2024, https://ico.org.uk/for-organisations/uk-gdpr-guidance-and-resources/designing-products-that-protect-privacy/childrens-code-design-guidance/protect-children-s-privacy-by-default/#:~:text=Privacy%20settings%20must%20be%20high,provide%20and%20never%20change%20them.
[35] eSafety Commissioner, 'Online Safety (Relevant Electronic Services—Class 1A and Class 1B Material) Industry Standard 2024,' Draft, accessed January 2024, https://www.esafety.gov.au/industry/codes/standards-consultation ; Onlinesafety.org.au, 'Schedule 1 – Social Media Services Online Safety Code (Class 1A and Class 1B Material)'
[36] 5Rights Foundation, 'Approaches to children's data protection'
[37] 5Rights Foundation, 'Pathways: How digital design puts children at risk,' July 2021, https://5rightsfoundation.com/uploads/Pathways-how-digital-design-puts-children-at-risk.pdf; 5Rights Foundation, 'Risky-by-design - case study 5: recommendation systems,' https://www.riskyby.design/recommendation-systems
[38] eSafety Commissioner, 'Position statement: recommender systems and algorithms,' 2022, https://www.esafety.gov.au/industry/tech-trends-and-challenges/recommender-systems-and-algorithms