

13 February 2024

Response to the Online Safety (Basic Online Safety Expectations) Amendment Determination 2023

This submission is made by the Australian Research Council (ARC) Centre of Excellence for the Digital Child, prepared by Associate Professor Anna Bunn (Curtin University), Dr Rys Farthing, Professor Tama Leaver, Professor Lelia Green, Professor Michael Dezuanni, Distinguished Professor Susan Danby and Dr Tara Roberson.

ABOUT THE ARC CENTRE OF EXCELLENCE FOR THE DIGITAL CHILD

The ARC Centre of Excellence for the Digital Child ('Centre'), funded by the Australian Research Council with AU\$34.9M over its seven-year life, is charged with leading national and global research, policy and practice to ensure that all Australian children are healthy, educated and connected in a rapidly expanding digital world. The Centre's research focuses on creating positive and safe digital experiences for children, aged from birth to eight years old. An internationally esteemed team of more than 40 interdisciplinary researchers with demonstrable expertise, diverse perspectives, and disciplinary expertise address the significant risks and opportunities of digital technologies in everyday lives of families and educators, including screen time, children's digital rights, e-privacy, commercialisation, digital technology innovation, relationships, health and wellbeing, sociality, education and learning, and digital play. The Centre involves six Australian universities, 14 international universities, and over 20 global partners, including Google and the Office of the eSafety Commissioner, along with national partners, including Early Childhood Australia, The Smith Family and SciTech. The Centre is therefore particularly well placed to respond to the BOSE Amendment and welcomes the opportunity to do so.

CONTACT INFORMATION

Associate Professor Anna Bunn: [REDACTED]

Distinguished Professor Susan Danby: [REDACTED]

Dr Tara Roberson: [REDACTED]

BOSE AMENDMENT: SUMMARY OF THE CENTRE'S POSITION

Broadly speaking, the Centre welcomes the reforms proposed in the BOSE Amendment. The speed and scale of development in the field of artificial intelligence ('AI') and other technologies with the potential for good and harm makes the proposed amendment timely. The explicit requirement for relevant services to take reasonable steps to ensure the best interests of the child is a primary consideration at all stages of design, development and deployment of services is particularly welcome. This approach aligns with Australia's obligations under the Convention on the Rights of the Child and with recent recommendations for reform to the *Privacy Act 1988* to introduce a requirement to consider the child's best interests in the context of personal information handling.

SUMMARY OF RECOMMENDATIONS

Overall, the Centre welcomes the BOSE Amendment but has several recommendations. A summary of the recommendations is set out below, followed by a more detailed response.

Best Interests

1. Make clear that services likely to be accessed by children, even if they are not targeted to children and even if they use age assurance mechanisms, will generally be expected to do more than services not likely to be accessed by children to demonstrate how they have considered the best interests of the child as a primary consideration and what they have done to ensure these interests have informed the design, deployment and operation of their services.
2. Amend BOSE paragraph 3(b) to: 'if a service or a component of a service (such as an online app or game) is likely to be accessed by, or is being used by, children (the *children's service*) ...'.
3. Ensure that further guidance is provided for services as to how to interpret and apply the best interests principle in practice.
4. Include non-limiting examples in the BOSE Determination of what reasonable steps are in the context of ensuring the best interests of the child are a primary consideration in the design and operation of any service, as per paragraph 7 (Detailed Response) below.
5. Require providers of services that are likely to be accessed by children to publish best interests assessments, as well as the steps they have taken to respond to any risks harms or other negative impacts identified through an assessment.

Hate Speech

6. Amend paragraph 6(3)(e)(i) to 'having processes for detecting and addressing hate speech'.
7. Amend the definition of hate speech to: 'For the purposes of paragraph 6(3)(i), hate speech is communication of any form by an end-user that attacks or uses pejorative or discriminatory language with reference to a person or a group on the basis of one or more identity factors, including (without limitation) race, ethnicity, disability, religious affiliation, caste, sexual orientation, sex, or gender identity; or on the basis of health status, immigrant status, asylum seeker or refugee status; or on the basis of age.'
8. Add a new paragraph in section 13(2) to require providers to have mechanisms that enable a person to report and make complaints about hate speech.

Deceptive Design/Dark Patterns

9. Insert a new section requiring providers to take reasonable steps to avoid deploying design features that have the effect of causing, or the potential to cause, harm to end-users.

10. Require services likely to be accessed by children to ensure that the best interests of the child are a primary consideration in the development or deployment of specific design choices in relation to the services.

Generative Intelligence

11. Amend paragraph 8A(3)(c) to: ‘ensuring that future training materials for generative artificial intelligence capabilities and models do not contain and are not trained on unlawful or harmful material and that every effort is made to ensure existing generative artificial tools do not create content based on unlawful or harmful material.’
12. Amend proposed paragraph 8A(3)(d) to read: ‘ensuring that generative artificial intelligence capabilities can detect and prevent the execution of prompts that generate unlawful or harmful material.’
13. Include examples of reasonable steps that could be take in relation to generative AI (as per paragraph 23 (Detailed Response) below).

Recommender Systems

14. Include non-limiting examples of recommender systems that are within scope of these provisions and ensure that these specifically include reference to friend recommender systems.
15. Amend proposed subsection 8B(1) to read: ‘If the service uses recommender systems, the provider of the service will take reasonable steps to consider end-user safety and incorporate safety measures in the design, implementation and maintenance of recommender systems on the service; and (without limiting the previous requirement) will take reasonable steps to ensure that the best interests of the child are a primary consideration in the design and operation of any recommender system that is likely to apply to or affect children.’
16. Include the ability for users to turn off recommender systems as an example of a reasonable step that could be taken in relation to such systems.

Proposed New Additional Expectations Regarding Systems and Elements

17. Include an overarching requirement on providers to ‘take reasonable steps to ensure that all systems and elements involved in the operation of the service can be used in a safe manner and to proactively minimise the extent to which those systems or elements of the service produce or promote material or activity that is unlawful or harmful.’
18. Require providers to consider the best interests of the child as a primary consideration in the development and deployment of all systems and elements, with specific examples being included of what is required here (in line with the suggestions in paragraph 7 (Detailed Response) below).

Age Assurance

19. Make clear that age assurance mechanisms must themselves be designed and used in the best interests of the child (or, collectively, of the children who will be impacted by them) and that this includes taking into account the risks they are intended to mitigate (eg. access to inappropriate content), and also risks inherent in the technology itself.
20. (Although not by way of amendment to the Determination) Commencing trials of age assurance technologies, involving the regulator, as recommended by the eSafety Commissioner in the ‘Roadmap to Age Verification’.
21. Include an additional requirement on providers to implement appropriate age assurance mechanisms for other levels of classified material (not only Class 2 material).

Enforcing Breaches of Terms of Use etc

22. Amend paragraph 14(1)(d) to: ‘standards of conduct for end-users (including in relation to material that may be posted using the service by end-users, if applicable), and policies and procedures in relation to the moderation of conduct and enforcement of those standards, which policies and procedures should reflect the best interests of the child’ and include additional guidance around this.

User Complaints

23. Amend proposed new section 15(2) as follows: ‘The provider of the service will ensure that the service has clear and readily identifiable mechanisms that enable any person ordinarily resident in Australia to report, and make confidential complaints about, breaches of the service’s terms of use and, where applicable, breaches of the service’s policies and procedures and standards of conduct mentioned in section 14.’

We note Reset Australia’s survey of over 1,000 young Australians, and follow up focus groups, which found that young people overwhelmingly want stronger protections in the digital environment, including stronger protections to keep them safe from abusive and distressing content in the digital world.¹ In light of this, and the upcoming review of the *Online Safety Act 2021*, we take this opportunity to express our strong support for amendments to the Act that will make BOSE expectations enforceable and will give the regulator power to hold providers to account.

Finally, we wish to emphasise the importance of consulting with children and young people on amendments to the BOSE Determination (if this has not already been done), as well as the ongoing impact of any amendments on them. This is necessary to fulfil Australia’s commitments under the *Convention on the Rights of the Child*², a commitment that has been contextualised by *General Comment No. 25 (2021) on Children’s Rights in the Digital Environment*, which provides that:

‘When developing legislation ... on children’s rights in relation to the digital environment, States parties should involve all children, listen to their needs and give due weight to their views.’³

This requires (as is also made clear in General Comment No. 25) ‘active engagement’ with children. The ARC Centre of Excellence for the Digital Child is well placed to facilitate consultation and welcome enquiries to that effect.

Our detailed response follows, below.

¹ Reset Australia, ‘Response to the Amending Online Safety (Basic Online Safety Expectations) Determination 2022 Consultation’, February 2024.

²CRC, art. 12 which provides that: ‘States Parties shall assure to the child who is capable of forming his or her own views the right to express those views freely in all matters affecting the child, the views of the child being given due weight in accordance with the age and maturity of the child.’

³ United Nations Committee on the Rights of the Child, *General Comment No. 25 (2021) on Children’s Rights in Relation to the Digital Environment*, CRC/C/GC/25 (2 March 2021), para. 17.

DETAILED RESPONSE

There are some aspects of the BOSE Amendment which the Centre feels need clarification or extension. We provide further details, below.

Best Interests of the Child

1. The proposed new subsection 6(2A) would require providers to take reasonable steps to ensure 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. We note that the phrase ‘any service used by, or accessible to, children’ means the provision is capable of broad application: i.e. it would apply to any service that children use or can access, regardless of whether the service is targeted at children and whether or not there are age assurance mechanisms in place designed to restrict access.
2. The Centre welcomes the inclusion of a broad obligation, given what the Consultation paper describes as the ‘unique vulnerabilities of children, and their particular susceptibility to online harms.’ However, we do have concerns that services not targeted to children, or that have age assurance mechanisms in place (even if the reality is that those services are likely to be accessed by children), could argue that, because of this, reasonable steps to ensure that the best interests of the child are a primary consideration require them to do little or nothing more than they already are doing. Accordingly, it should be clarified that, while the requirement to take reasonable steps to ensure best interests of the child are a primary consideration applies to *all* services used by or accessible to children, services ‘likely to be accessed by children’ (even if they are not targeted to children and even if they use age assurance mechanisms) will generally be expected to do more than services not likely to be accessed by children to demonstrate how they have considered the best interests of the child as a primary consideration and what they have done to ensure these interests have informed the design and deployment of their services.
3. We suggest amending existing paragraph 3(b) to replace the words: ‘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**)’ with ‘if a service or a component of a service (such as an online app or game) is likely to be accessed by, or is being used by, children (the **children’s service**)’. This amendment promotes regulatory consistency by bringing this requirement more in line with the proposal to introduce a Children’s Online Privacy Code that would require (if enacted as recommended) high privacy-by-default settings for services that are likely to be accessed by children (regardless of whether they are targeted at them, or actually used by them).⁴
4. Although the inclusion of subsection 6(2A) sets an overarching obligation on providers, it would help to keep front and centre the requirement to consider the child’s best interests as a primary consideration if reference to best interests were also included in examples of reasonable steps relating to generative artificial intelligence (AI) and recommender systems. Specific suggestions on this are set out below, under the headings ‘Generative AI’ and ‘Recommender Systems’.
5. Provide guidance on interpreting the requirement to consider best interests as a primary consideration. Make clear (as per the Consultation Paper) that the requirement is to consider

⁴ Attorney-General (Cth), ‘Privacy Act Review Report 2022’, 10, Proposal 16.5.

the best interests of the child generally (rather than a particular child) and provide guidance on what it means to take children's interests into account as a primary consideration and what, in practice, this involves. Specifically, guidance should make it very clear that simply implementing age assurance mechanisms does not remove the necessity to further consider children's best interests. In addition, make clear that the requirement to take reasonable steps will generally require providers to do more than simply making available user controls to support safe online interactions. This recommendation is made because, as noted in our Manifesto for a Better Children's Internet:

a significant amount of parent or carer labour is associated with the management of children's digital experiences and the technology companies rely on this labour as part of their justification for making products and experiences available to children. They also rely on this parent labour to significantly reduce the costs associated with direct moderation and more advanced technological solutions and design features that may make children's experiences more private, safe, and pleasant.⁵

6. Guidance provided by Reset Australia on interpreting the best interest of the child in the context of targeting is instructive.⁶ Guidance provided by the UK Information Commissioner's Office is also useful, particularly in terms of highlighting common data protection activities that can impact children's rights.⁷
7. Additionally, give consideration to including in the BOSE Determination examples of what 'reasonable steps' are in the context of ensuring that the child's best interests are a primary consideration in the design and operation of any service. We suggest the following, non-limiting examples, be included:
 - (a) That providers of services used by, accessible to or likely to be accessed by children, consult with appropriate experts (eg. in age-appropriate/child-safe design) when designing services and/or any age assurance mechanisms that will be used when accessing these services;
 - (b) That providers conduct 'best interests assessments' with the meaningful involvement of experts (as per above) as well as children and young people, when developing new services or components that are likely to be accessed by children (an example of what a best interests assessment might look like in the context of targeting has been developed by Reset Australia and guidance materials along these lines could be developed to support providers⁸);
 - (c) That providers conduct regular 'best interests assessments' with the meaningful involvement of experts (as per above) as well as children and young people, during the lifetime of services or components of services that are likely to be accessed by children;

⁵ ARC Centre of Excellence for the Digital Child, 'Manifesto for a Better Children's Internet' (2023) Digital Child Working Paper Series, 2023-11, 71 [ChildrensInternet-Interactive-1.pdf \(digitalchild.org.au\)](https://digitalchild.org.au/ChildrensInternet-Interactive-1.pdf).

⁶ Reset Australia, 'Best Interests & Targeting: Implementing the Privacy Act Review to Advance Children's Rights', Policy Briefing, January 2024.

⁷ Information Commissioner's Office (UK), *UK GDPR Guidance and Resources, Best Interests of the Child Self-Assessment* <https://ico.org.uk/for-organisations/uk-gdpr-guidance-and-resources/childrens-information/childrens-code-guidance-and-resources/best-interests-self-assessment/step-2-identifying-impacts-on-childrens-rights/common-data-processing-activities-that-can-impact-children-s-rights/>.

⁸ Reset Australia, 'Best Interests & Targeting: Implementing the Privacy Act Review to Advance Children's Rights', Policy Briefing, January 2024.

- (d) That providers take steps to respond to risks, harms and other potentially negative effects of the design or operation of the service or any component of it that are identified in any best interests assessment;
 - (e) For all services likely to be accessed by children: switching ‘geolocation’ data off by default; alerting children when their location is tracked or visible to others; and reverting to default ‘off’ settings for geolocation data after each use. (This reflects provisions in the UK Age-Appropriate Design Code that have been included because, according to the UK Information Commissioner, ‘the ability to ascertain or track the physical location of a child carries with it the risk that the data could be misused to compromise the physical safety of that child. In short it can make children vulnerable to risks such as abduction, physical and mental abuse, sexual abuse and trafficking’⁹).
8. These provisions could be strengthened by including a requirement for providers of services likely to be accessed by children to publish their best interests assessments, as well as the steps they have taken to respond to any risks, harms or other negative impacts identified in an assessment. This addition increases transparency and provides an opportunity for service providers to share their methods and processes.

Hate Speech

9. Proposed new paragraph 6(3)(i) requires providers to have ‘processes for detecting and addressing hate speech which breaches a service’s terms of use and, where applicable, breaches a service’s policies and procedures and standards of conduct mention in section 14.’ Proposed new subsection 6(4) would then insert a definition of ‘hate speech’ for the purposes of paragraph 6(3)(e) which definition also references speech that is contrary to the terms of use and policies, procedures and standards of providers.
10. This is problematic. First, section 14 does not require providers to include terms and conditions or policies and procedures dealing with hate speech per se. It requires providers to have policies and procedures in relation to the safety of end-users and for dealing with reports about complaints mentioned in sections 13 or 15 (that are directed at certain types of conduct which do not necessarily include the type of communications envisaged under the proposed new subsection 6(4)). Therefore, those providers who do *not* already have terms of use or policies, procedures or standards dealing with hate speech will not be required to have processes for detecting and addressing hate speech, which is counter intuitive.
11. Where providers do have terms and conditions, policies or standards directed at ‘hate speech’, the scope of these terms (and what is included as ‘hate speech’ or is regulated as unacceptable speech or conduct) varies between providers.¹⁰ This means the regulation of ‘hate speech’ would be dependent on what types of speech are regulated by the provider itself, leading to inconsistency and subjectivity. Given that, as noted in the Consultation

⁹ Information Commissioner’s Office (UK), *Age-appropriate Design: A Code of Practice for Online Services’ Reference*, Standard 10 <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/>.

¹⁰ Mika Hietanen and Johan Eddebo, ‘Towards a Definition of Hate Speech-With a Focus on Online Contexts’ (2020) 47(4) *Journal of Communication Inquiry* 440, 448 <https://doi.org/10.1177/01968599221124309>.

paper, many Australian adults are targeted by online hate, which ‘disproportionately impacts many groups’, this is unacceptable.

12. For these reasons, proposed paragraph 6(3)(e)(i) should be amended as follows:

having processes for detecting and addressing hate speech ~~which breaches a service’s terms of use and, where applicable, breaches a service’s policies and procedures and standards of conduct mentioned in section 14.~~

13. Amend the definition of hate speech in subsection 6(4). We suggest that the definition be framed objectively (i.e. not be dependent on what is covered by terms, conditions, policies, standards etc).

14. Fixing an appropriate definition may require further consultation and is complex.¹¹ We suggest that an objective definition could be considered by adapting the United Nations definition of hate speech.¹² A suggested amendment to the definition (changes indicated) might be:

For the purposes of paragraph 6(3)(i), hate speech is communication of any form by an end-user that ~~breaches a service’s terms of use and, where applicable, breaches a service’s policies and procedures or standards of conduct mentioned in section 14, and~~ can include communication which attacks or uses pejorative or discriminatory language with reference to a person or a group on the basis of one or more identity factors, including (without limitation) race, ethnicity, disability, religious affiliation, caste, sexual orientation, sex, or gender identity; or on the basis of health status, immigrant status, asylum seeker or refugee status; or on the basis of age.

15. Although there are risks in adopting a unitary definition of ‘hate speech’, a requirement for providers to have processes for detecting and addressing speech that falls within the definition we have proposed does not require providers to take any particular action in response that speech. So, for example, speech that fits within the definition does not necessarily need to be censored and ultimately providers can consider context and other matters in determining an appropriate response.

16. Add a new paragraph in section 13(2) to require providers to have mechanisms to enable a person to report and make complaints about hate speech.

Manipulative Design Features

17. Certain design features have the capacity to manipulate users, impact their safety and undermine their best interests. Children, in particular, may be more susceptible to manipulative designs.¹³ Manipulative design features (also referred to as ‘dark patterns’ or

¹¹ Mika Hietanen and Johan Eddebo, ‘Towards a Definition of Hate Speech-With a Focus on Online Contexts’ (2020) 47(4) *Journal of Communication Inquiry*, <https://doi.org/10.1177/01968599221124309>.

¹² United Nations, ‘Understanding Hate Speech’, *Hate Speech* <https://www.un.org/en/hate-speech/understanding-hate-speech/what-is-hate-speech>.

¹³ Jenny Radesky, Alexis Hiniker, Caroline McLaren et al, ‘Prevalence and Characteristics of Manipulating Design’ (2022) 5(6) *JAMA Netw Open*, 2 doi:10.1001/jamanetworkopen.2022.17641; Office of Fair Trading (UK), *Children’s Online Games: Report and Consultation*, Report No. 1506 (2013), 13 <https://assets.publishing.service.gov.uk/media/53330c4de5274a5660000005/oft1506.pdf>.

‘deceptive design’) can, among other things, ‘nudge’ children to share more personal information than is necessary for the function of the service; prompt children to spend money; and encourage addiction.¹⁴ We therefore propose a new section be inserted into the BOSE Determination requiring providers to take reasonable steps to avoid deploying design features that have the effect of causing or potential to cause harm to end-users.

18. Require services likely to be accessed by children to ensure that the best interests of the child are a primary consideration in the development or deployment of specific design choices. Although there is already an overarching obligation on providers to ensure the best interests of the child is a primary consideration in the development and operation of services, specific reference to design features focusses attention on the particular problems presented by manipulative design.

Generative Artificial Intelligence

19. We welcome specific provisions regulating generative artificial intelligence (AI).
20. While we commend the intent of paragraph 8(3)(c), one of the biggest challenges with the models that underpin many generative intelligence tools (a common example of which is large language models, or LLMs) is that it is almost impossible retrospectively to know what material they were trained on.¹⁵ Thus, the current wording would make almost all LLMs/generative AI tools unusable.

21. We recommend rephrasing proposed paragraph 8A(3)(c) to:

‘ensuring that future training materials for generative artificial intelligence capabilities and models do not contain and are not trained on unlawful or harmful material and that every effort is made to ensure existing generative artificial tools do not create content based on unlawful or harmful material.’

22. In terms of examples of reasonable steps that could be taken in relation to AI, we suggest that paragraph 8A(3)(d) is amended to read (changes indicated):

‘ensuring that generative artificial intelligence capabilities can detect and prevent the execution of prompts that generate unlawful or harmful material.’

23. We also suggest that additional examples of reasonable steps that could be taken are added, as follows:

- (a) Enabling end-users to make complaints and address enquiries about the role that AI may play in presenting material or activity on the service that is unlawful or harmful;

¹⁴ Carla Sousa and Ana Oliveira, ‘The Dark Side of Fun: Understanding Dark Patterns and Literacy Needs in Early Childhood Mobile Gaming’ (2023) 17(1) *Proceedings of the 17th European Conference on Games Based Learning* <https://doi.org/10.34190/ecgbl.17.1.1656>; Jenny Radesky, Alexis Hiniker, Caroline McLaren et al, ‘Prevalence and Characteristics of Manipulating Design’ (2022) 5(6) *JAMA Netw Open*, 2 doi:10.1001/jamanetworkopen.2022.17641; Office of Fair Trading (UK), *Children’s Online Games: Report and Consultation*, Report No. 1506 (2013), 13 <https://assets.publishing.service.gov.uk/media/53330c4de5274a5660000005/oft1506.pdf>.

¹⁵ Emily M Bender, et al, ‘On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? 🦜’ (2021) *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency, ACM, 2021*, 610 <https://doi.org/10.1145/3442188.3445922>.

- (b) Ensuring that the design, implementation and maintenance of AI capabilities complies with Australia’s AI Ethics principles (as updated from time to time) and with any Australian standards that may be adopted in relation to the use of AI generally or relating to the particular AI system or capability being developed or deployed;
- (c) Ensuring regular independent audits of the function of generative AI systems.

Recommender Systems

24. Non-limiting examples of recommender systems that are within the scope of these provisions should be included in the Determination or accompanying guidance. These examples should specifically include friend recommender systems, given that it has been reported that one Meta’s “People You May Know” algorithm ‘was known among employees to connect child users with potential predators’.¹⁶

25. We also propose adding to subsection 8B(1), as follows (change indicated):

If the service uses recommender systems, the provider of the service will take reasonable steps to consider end-user safety and incorporate safety measures in the design, implementation and maintenance of recommender systems on the service; and (without limiting the previous requirement) will take reasonable steps to ensure that the best interests of the child are a primary consideration in the design and operation of any recommender system that is likely to apply to or affect children.

26. Examples of reasonable steps that could be taken in relation to recommender systems should include the ability for users to ‘turn off’ these systems.

New Section: Additional Expectations Regarding Systems and Elements

27. We welcome provisions addressing the risks posed by Generative AI and recommender systems. However, these are not the only systems that pose risks to users and focussing on these, to the exclusion of others, fails to capitalise on the opportunity to better protect Australians and future-proof the Determination.

28. Therefore, we suggest that the BOSE Amendment includes an overarching requirement on providers to:

‘take reasonable steps to ensure that all systems and elements involved in the operation of the service can be used in a safe manner and to proactively minimise the extent to which those systems or elements of the service produce or promote material or activity that is unlawful or harmful.’

29. Provide non-exhaustive examples of systems and elements that are (at present) of particular concern. These would include generative AI and recommender systems as well other systems (such as content moderation systems, user control systems, advertising approval systems, advertising management systems, and others). For clarity, this section should not

¹⁶ Katherine Blunt and Jeff Horwitz, ‘Children on Instagram and Facebook Were Frequent Targets of Sexual Harassment, State Says’, *The Wall Street Journal* (online, 17 January, 2024) <https://www.wsj.com/tech/children-on-instagram-and-facebook-were-frequent-targets-of-sexual-harassment-state-says-68401b07>.

replace the specific provisions on generative AI and recommender systems, although consideration could be given to including more detailed provisions in relation to some other systems.

30. Require providers to consider the best interests of the child as a primary consideration in the development and deployment of all systems and elements; with specific examples being included of what is required here (in line with the suggestions in paragraph 7 above).

Age Assurance for Class 2 Materials

31. We note that the eSafety Commissioner concluded that age assurance technologies come with their own ‘privacy, security, effectiveness and implementation issues’ and that the market for age assurance technologies is ‘immature, but developing’. We therefore recommend Including in the BOSE Amendment or accompanying guidance a reminder to providers that age assurance mechanisms must themselves be designed and used in the best interests of the child (or, collectively, of the children who will be impacted by them): this includes taking into account not only the risks they are intended to mitigate (eg access to inappropriate content) but also risks inherent in the technology itself.
32. Although not by way of an amendment to the Determination, we are of the view that trials of age assurance technologies, as recommended by the eSafety Commissioner in the ‘Roadmap to Age Verification’,¹⁷ should take place and should involve the regulator, in order to increase community trust in the process. These trials should not be put on hold until the outcome of the class 2 industry codes process is known, given the high likelihood that this process will not result in the development of satisfactory codes across the industry without regulator involvement (as per the Class 1 material codes).

Additional Provision: Age Assurance for Other Content

33. We suggest that providers are required to implement appropriate age assurance mechanisms for all levels of classified material (not only Class 2 material).

Enforcing Breaches of Terms of Use etc

34. While we welcome proactive steps to enforce penalties for breaches of terms of use etc, we should also note that many breaches will be perpetrated by children and young people themselves, commonly teenagers who push boundaries in order to ‘shock’ or ‘scare’ vulnerable children. Service providers should, in the interest of the child, have a policy that covers this. We suggest an amendment to paragraph 14(1)(d) as follows (amendment indicated):

standards of conduct for end-users (including in relation to material that may be posted using the service by end-users, if applicable), and policies and procedures in relation to the moderation of conduct and enforcement of those standards, which policies and procedures should reflect the best interests of the child.

¹⁷ eSafety Commissioner, ‘Roadmap for Age Verification and Complementary Measures to Prevent and Mitigate Harms to Children from Online Pornography’, Report (March 2023)
https://www.esafety.gov.au/sites/default/files/2023-08/Roadmap-for-age-verification_2.pdf.

35. Guidance could then be issued such that providers should consider the fact that many of those in breach of terms and conditions, policies and procedures will themselves be children or young people and that their interests should also be factored in when determining appropriate enforcement action and opportunities for educative action.

User Complaints

36. To reassure users that their identity will not be revealed should a complaint be made, we suggest amending proposed new section 15(2) as follows (amendment highlighted):

The provider of the service will ensure that the service has clear and readily identifiable mechanisms that enable any person ordinarily resident in Australia to report, and make confidential complaints about, breaches of the service's terms of use and, where applicable, breaches of the service's policies and procedures and standards of conduct mentioned in section 14.