Online Safety Amendment (Digital Duty of Care) Bill 2026

Submission to the Environment and Communications Legislation Committee

30 October 2026

www.humanrights.gov.au

Australian Human Rights Commission

Contents

Overview **4**
Recommendations 6
1 Beneficial measures within the Bill **7**
Safe online environment 7
Content exclusions 7
Seriously harmful material and conduct 8
Best interests of children 10
Synthetic nude material 11
Adult-child bifurcated model 12
Improvements for researchers 13
Scope of captured services 15
Legislating Codes and Standards and the Basic Online Safety Expectations 16
eSafety Guidelines 17
Transparency and publication 17
Complaints and dispute processes 17
Link deletion and faster removal 18
Design features definition 19
2 Improvements needed **20**
Harmful to children definition 20
User empowerment tools 24
Risk assessments 25
Political neutrality 25
Moderation and HRIA 25
3 One-year statutory review **26**
Endnotes **27**

About the Australian Human Rights Commission

Our vision is an Australian society where human rights are respected, promoted and protected and where every person is equal in dignity and rights.

The Commission's key functions include:

- Access to justice: We help people to resolve complaints of discrimination and human rights breaches through our investigation and conciliation services.
- Fairer laws, policies and practices: We review existing and proposed laws, policies and practices and provide expert advice on how they can better protect people's human rights. We help organisations to protect human rights in their work. We publish reports on human rights problems and how to fix them.
- Education and understanding: We promote understanding, acceptance and public discussion of human rights. We deliver workplace and community human rights education and training.
- Compliance: We are the regulator for positive duty laws requiring employers and others to address sexual harassment, sex discrimination and other unlawful conduct.

Australian Human Rights Commission www.humanrights.gov.au

ABN 47 996 232 602

GPO Box 5218 Sydney NSW 2001

For general enquiries, call us on 1300 369 711.
For complaints, call us on 1300 656 419.
For TTY, call 1800 620 241.

Overview

1. We have consistently called for a legislative digital duty of care (DDC) to protect people from online harm. Online services should take responsibility for identifying and addressing foreseeable risks, rather than leaving individuals to manage those risks. A proactive, systems-based duty offers an important opportunity to improve online safety while preserving the benefits of digital participation.
2. The case for a DDC has been made repeatedly through parliamentary committees and reports:
   - The Senate Legal and Constitutional Affairs References Committee recommended a legislated duty of care on social media platforms in its 2018 cyberbullying inquiry.[1]
   - The Select Committee on Social Media and Online Safety also recommended examination of a statutory duty, although this was a recommendation for further consideration (rather than immediate introduction).[2]
   - The Joint Select Committee on Social Media and Australian Society recommended a single, overarching statutory duty supported by risk assessments and mitigation plans in 2024.[3]
   - The Senate Environment and Communications Legislation Committee called for a DDC in its inquiry into the *Social Media Minimum Age Bill* (Cth) in 2024.[4]
   - The Senate Environment and Communications References Committee renewed that call in its 2025 inquiry into the Internet Search Engine Services Online Safety Code and under-16 Social Media Ban.[5]
3. Independent reviews have reinforced the need for proactive obligations. The 2018 Briggs review recommended legislated obligations on industry to detect, remove and deter illegal and harmful content.[6] The 2024 Rickard statutory review of the *Online Safety Act 2021*(Cth) (OS Act) recommended a single, overarching DDC encompassing due diligence, safety-by-design, risk assessment and risk mitigation.[7]
4. We have consistently supported this direction including through submissions to the:
   - Statutory review of the OS Act.[8]
   - Senate Standing Committees on Environment and Communications on the *Online Safety Amendment (Social Media Minimum Age) Bill 2024* and accompanying public explainer.[9]

- Human Rights Council Advisory Committee inquiry into Technology-Facilitated Gender-Based Violence.[10]
- Senate Environment and Communications References Committee on the Search Engine Safety Code and Social Media Ban.[11]
- Department of Infrastructure, Transport, Regional Development, Communications, Sport and the Arts on the DDC Issues Paper in July 2026 and then the *Exposure Draft of the Online Safety Amendment (Digital Duty of Care) Bill 2026* (Cth) (Exposure Draft) in September 2026.[12]

5. Across this work, we have consistently emphasised that preventative regulation must address systemic risks while ensuring regulatory measures do not unduly restrict freedom of expression, privacy, equality and children's rights. Effective legislation should protect people from harm while enabling them to seek information, express views, participate in public debate and maintain connections online.

6. The *Online Safety Amendment (Digital Duty of Care) Bill 2026* (Cth) (Bill) contains important improvements from the Exposure Draft. We welcome the excluded content carveout, narrower definitions and removal of the Minister's discretionary power to expand the harm categories. We also support:

   - recognition of children's best interests
   - broad coverage of online services including generative AI
   - stronger research and transparency arrangements and
   - accessible complaints processes.

7. However, support for the DDC does not diminish the importance of getting its details right. Broad or uncertain harm categories may encourage services to restrict lawful material rather than assess its context. The excluded content carveout is a valuable safeguard, but does not remove the need for clear underlying definitions.

8. Separate protections for adults and children can better reflect their different needs. However, age assurance must be necessary, proportionate and privacy-preserving.

9. The framework should also expressly incorporate human rights into risk assessment and mitigation. Services are likely to expand their use of AI-informed moderation tools to meet the DDC and should be required to conduct human rights impact assessments before deploying or significantly changing those tools. Key policy settings should be established in primary legislation, and the framework must remain politically neutral in both its terms and application.

10. Finally, the Bill should require a statutory review one year after the DDC commences. The review should examine the operation of the DDC including,

inter alia, whether the harm categories should expressly include hate speech and whether people harmed by a breach of the legislation should have a private right of action. These issues require careful consideration of the full range of human rights involved. With the improvements identified in this submission, the DDC can provide a stronger foundation for a safer, rights-respecting digital environment.

Recommendations

11. We support the objective of the Bill as it will provide an important foundation for improving online safety. This submission outlines a number of amendments that are necessary to ensure the framework appropriately protects human rights including (but not limited to) the recommendations below.

Recommendations
1. The Australian Government delete 'Australian' from s 3(b) of the Bill and adds '… as provided for in the Convention on the Rights of the Child'.
2. The Australian Government should remove the 'encourages seriously harmful practices' limb of s 25E(1)(d) or clearly define its scope in primary legislation to reduce the risk of unjustified restrictions on lawful expression and access to information.
3. The Australian Government should amend s 25E(1)(f) to replace 'may cause' with 'is reasonably likely to cause', to provide a clearer threshold and reduce the risk of overreach in content moderation.
4. The Australian Government should establish key user empowerment tools in primary legislation, subject to parliamentary scrutiny. If these tools are instead prescribed through ministerial rules, the Government should ensure public stakeholder consultation and adequate opportunities for public scrutiny of proposed rules before they are made.
5. The Australian Government should amend the Bill to require a statutory review one year after the digital duty of care commences, including consideration of, inter alia, whether ss 25D and 25E should expressly include hate speech and whether people harmed by a breach of the duty should have a private right of action.

1 Beneficial measures within the Bill

This section supports the DDC's proactive approach to online safety and highlights clearer definitions, public-interest exclusions and children's best interests. We also understand the need for separate protections for adults and children, supported by privacy safeguards.

12. We support the objective of the Bill as it will provide an important foundation for improving online safety. However, a number of amendments (including, but not limited to, those listed in this submission) are necessary to ensure the framework appropriately protects the full suite of human rights.

Safe online environment

13. The Bill requires a person responsible for online services to take effective measures to ensure a 'safe online environment'.[13] This differs for adults and children, with adults protected from 'seriously harmful material and conduct' and children protected from a broader range of harmful material, conduct and design-related harms.[14]

14. However, it is concerning that the Bill no longer expressly requires privacy impacts to be considered when determining the measures necessary to comply with the DDC.[15] Given the potential privacy implications of age assurance (as discussed below), it is necessary to provide guidance in the Explanatory Memorandum to clarify that s 26A of the Bill requires consideration of human rights more generally – which would include the right to privacy.

Content exclusions

15. A notable improvement from the Exposure Draft, has been the inclusion of s 25F which sets out 'excluded content' which is not captured as 'seriously harmful' material under s 25D or content that is 'harmful to children' under s 25E. We previously raised concerns about potential censorship without such a carveout.[16] The exclusions in s 25F are a substantial improvement and provide important safeguards for freedom of expression.

16. However, additional clarity should be provided regarding political communication and matters of public debate. Greater certainty should also be provided regarding the operation of the public-interest exception. This is particularly important where moderation decisions will often be made by platforms rather than courts.

17. The content exclusions also help to protect the right to seek, receive and impart information, as contained in art 19 of International Covenant on Civil and Political Rights (ICCPR). The breadth of the carveout in relation to legacy and

new media is welcomed. This is important because until recently most Australians received information about current affairs through TV, radio or newspaper.[17] While legacy media remain the most popular sources of news (37%), social media platforms (26%) have now overtaken online news sites (23%) for people in Australia.[18] It is important that these carveouts remain broad enough to protect both legacy and new media.

Seriously harmful material and conduct

18. Given the DDC may influence content moderation practices or lead to a restriction of information online – clarity and precision are essential for the definitions used. We have previously criticised the definitions in respect of 'seriously harmful material and conduct'.[19] However, the amendments made to the Bill appear to rightly narrow the definitions to be both more workable and better calibrated to protect freedom of expression (when read in conjunction with the carve out for excluded material in s 25F).

19. Notably we welcome the inclusions of:

- non-consensual deepfake sexual abuse material in s 25D(1)(d)
- posting of, or threatening to post, an intimate image without consent in s 25D(1)(e) and
- gender-based violence in s 25D(1)(h).

20. CEDAW arts 2 and 5 and General Recommendation No. 35 makes clear that States must prevent gender-based violence against women, not simply respond after it occurs. It calls for action to address the prejudices, stereotypes and attitudes that help sustain such violence, including through education and cooperation with media and digital platforms.[20] This provides a strong basis for preventative online safety measures that address these underlying causes.

21. However, it does not automatically justify restricting all expression - as any restriction must still be clearly defined in law, necessary and proportionate under art 19(3) ICCPR.[21] This is particularly relevant when considering the below points.

Menacing or harassing an individual

22. The amendment of s 25D(1)(i) is a welcome clarification. The revised test focuses on whether serious harm is 'reasonably likely' - rather than whether it was 'intended'.[22] Clarifying that 'serious harm' is consistent with the definition in s 5 of the *Online Safety Act 2021* (Cth) (OS Act) also provides greater certainty about the provision's scope.[23] The narrowing of the provision makes s 25D(1)(i) more workable and reduces the risk of unnecessarily infringing on freedom of expression.

23. Article 19(2) ICCPR protects the freedom to seek, receive and impart information and ideas of all kinds. Protection extends even to expression that may be regarded as offensive.[24] However the right is not absolute and may be subject to restrictions – but only if those restrictions are provided by law and are necessary and proportionate.[25]

24. The requirement that a restriction be provided by law demands more than simply including it in the Bill. General Comment No. 34 requires laws restricting expression to be sufficiently precise.[26] In this context, 'menacing' and 'harassing' identifies the threatening or abusive character of the conduct, rather than operating as mere labels for unpopular viewpoints. Read alongside the requirement that serious harm be reasonably likely, these terms provide a sufficiently precise legal basis for a restriction on expression.

25. Even where restrictions are determined to be legal, they must also meet strict tests of necessity and proportionality – requiring the restriction to:
    - pursue a legitimate aim, including respecting the rights of others
    - be proportional to that aim and
    - be no more restrictive than is necessary to achieve that aim.[27]

26. Section 25D(1)(i) pursues the legitimate aim of protecting the rights of others by addressing menacing or harassing material or conduct reasonably likely to cause serious harm.[28] Its inclusion in a preventative duty is justified by the limitations of the previous complaints-based approach under the OS Act – which responded after harm had occurred.[29] Its proportionality is supported by the combined requirements of menacing or harassing conduct and a reasonable likelihood of serious harm, together with the public-interest exclusions in s 25F.

27. These features support the provision's compatibility with art 19(3) ICCPR, provided that measures taken to implement it address demonstrated risks and use the least intrusive effective means, rather than automatically removing all material within the category.

28. At the same time, the provision may also fail to capture the cumulative impact of harassing or menacing content directed at an individual.[30] Individual posts or messages may not meet the serious harm threshold when assessed separately, even though their combined effect may be seriously harmful. The drafting should clarify that the likelihood of serious harm may also be assessed cumulatively, rather than by considering each item in isolation.

29. International human rights law recognises that protecting the rights of others includes protecting individuals as members of a community, and that threats and incitement may be directed against groups rather than named individuals.[31] Committee on the Elimination of Racial Discrimination General Recommendation No. 35 stresses contextual assessment, differentiated responses and protection for debate about matters of public interest.[32] These

will all be necessary considerations for online services as they take measures to meet their obligations under the DDC.

30. However, limiting such conduct to an 'individual' does not reflect the way identity-based hostility operates online. Evidence given during the Royal Commission into Antisemitism and Social Cohesion demonstrated that antisemitism may operate against Jewish Australians as a group.[33] Similarly our Equal Identities report found that hostility operates systemically – targeting the group even where it didn't name an identifiable person.[34] An individual-only provision may miss cumulative conduct directed at a protected group, even where no single person is named.

Encouraging, promoting, inciting or instructing criminal offences

31. The narrowing of this category of serious harm in s 25D(1)(m) assists in ensuring that restrictions on the right to protest are no broader than necessary to protect people from conduct that risks serious harm.

32. Unlike the Exposure Draft, the revised provision is confined to the commission of a serious Commonwealth offence that is reasonably likely to cause serious harm.[35] The Commonwealth-only limitation, three-year imprisonment threshold and additional serious harm requirement reduces the range of protest-related activity potentially captured (particularly activity criminalised solely under State or Territory law).

33. Notably section 25D(1)(m) may capture some forms of hate speech, including material encouraging violence against protected groups where there is overlap with the offences in Division 80 of the *Criminal Code Act 1995* (Cth).

Removal of Minister's power

34. We support the removal of the Minister's discretionary power to add categories of seriously harmful material and conduct.[36] Given the potential restrictions on lawful expression, access to information and rights to protest, any expansion of the scope of seriously harmful material and conduct should be made through primary legislation and subject to full parliamentary scrutiny. This provides greater certainty about the scope of the DDC and ensures that proposed restrictions can be assessed for their necessity, proportionality and human rights implications.

Best interests of children

35. We support the express inclusion of protecting, and acting in the best interests of, children online in the objects section of the Bill.[37] The best interests of children is a key principle in the Convention on the Rights of the Child, and reference could be made to the Convention within the Bill. .

36. The inclusion of 'Australian' may have the unintentional impact of limiting the objective of the DDC to just those children who are Australian citizens – rather than all children who are in Australia. This should be amended to ensure all those under 18 are protected.

Recommendation: The Australian Government delete 'Australian' from s 3(b) of the Bill and adds '… as provided for in the Convention on the Rights of the Child'.

37. Article 3(1) of the Convention on the Rights of the Child requires that the best interests of the child be a primary consideration in all actions concerning children. Operationalising the best interests of the child requires online services to place children's rights and wellbeing at the centre of risk assessment and mitigation decisions. It does not require online services to eliminate all risk or to impose child-protective restrictions on adults by default. Rather, it means that services should make proportionate decisions about foreseeable risks – while preserving young people's ability to participate, learn, seek information and express themselves online.

38. Arguments that the best interests principle is too uncertain to be implemented should not be accepted. Australian regulatory design is already moving in the same direction. The proposed Children's Online Privacy Code is anchored in the best interests of the child and requires regulated services to consider children's best interests when designing and operating services that collect, use or disclose children's personal information.[38] That parallel framework demonstrates experience with practical implementation in digital regulation.

39. Importantly, the best interests principle is not a matter of simply doing no harm – it is about ensuring children's interests are at the centre of all decisions made under the DDC. We are concerned that online service may not interpret this provision in the spirit of the Convention on the Rights of the Child.

Synthetic nude material

40. Deepfake sexual abuse is a form of image-based sexual violence where sexually explicit content is made using someone's image without their consent.[39] Women and girls are overwhelmingly the targets of deepfake sexual abuse,[40] with eSafety Commissioner, Julie Inman Grant, noting there has been a 550% year-on-year increase in global material since 2019.[41] In the first nine months of 2023 alone almost 150,000 sexualised deepfakes were uploaded - more than all previous years combined.[42]

41. The exponential rise of deepfake sexual abuse and synthetic explicit material is a form of technology-facilitated gender-based violence which directly undermines arts 5-7 of the *Convention on the Elimination of Discrimination Against Women* (CEDAW).

42. While the impact to women who are depicted in deepfake sexual abuse material is very real – so too is the impact on gender equality more broadly when people can generate any synthetic nude material (regardless of whether it depicts a real person). These new forms of online harm overwhelmingly target women and girls and reinforce harmful misogynistic stereotypes and objectification.

43. Furthermore, weaponising synthetic material against women functions as a tool of cyber-harassment that creates a chilling effect - severely restricting their safe and equal participation in political, professional and public life.[43] In alignment with CEDAW General Recommendation No. 35, State parties have a clear obligation to actively regulate these digital spaces - reinforcing the urgent need to strengthen the Office of the eSafety Commissioner's (eSafety) powers to remove the tools driving this systemic harm.[44]

44. We support strengthening eSafety's powers to remove and penalise apps or websites that are designed or used predominantly to generate fake nude material – even where a real person is not depicted.[45]

Threshold

45. Yet the threshold is narrow because it only applies to apps and websites whose predominant purpose or use is to generate synthetic nude material. This means that generative AI tools that have other purposes will not be captured - even where they are used to create synthetics nudes. This is an issue with s 86A that Parliament should consider further.

Adult-child bifurcated model

46. We agree that the DDC should be bifurcated because the online experiences of children and adults are different. Australia has adopted a bifurcated approach to digital regulation across several regimes.[46] This bifurcated model will avoid the application of restrictive child-settings to adults.

47. Yet the shift toward a dual digital experience affects how young people encounter information and form views. Earlier regulatory models, such as film classification, operated in a context where content could be clearly categorised and limited. In contrast, the online environment is continuous and embedded in everyday life.

48. A bifurcated model should recognise that children's abilities and independence develop over time.[47] Applying the same restrictions to all children risks overlooking differences in their maturity and understanding.[48] In this context, a segmented online experience risks producing a narrow or mediated representation of the world for younger users. While restricting access to harmful content is appropriate, overly rigid filters limit the opportunity for gradual exposure and critical engagement. This is particularly relevant in

relation to civic and political information, where exposure to different perspectives supports informed participation and the development of independent views.

49. The operation of a bifurcated approach will also increase the use of age-assurance technologies, with serious implications for privacy, anonymity and data security. While we recognise that this means age assurance will necessarily be increasingly used, it does not remove the importance of ensuring that age assurance is used in a manner that is consistent with human rights.

Age assurance

50. A difficulty of the bifurcated model is that it relies upon age assurance technologies, which have problems with accuracy for women and people from culturally and linguistically diverse groups.[49] There is also growing evidence that the age assurance measures being implemented in Australia are not effective.[50]

51. The increasing reliance on age assurance warrants scrutiny. While there are few workable alternatives to enable a bifurcated experience, the growing dependence on age assurance raises significant concerns – especially around privacy under art 17 ICCPR.

52. We recognise that some have argued that extending the same child-proof protections to adults would avoid the privacy risks associated with age assurance.[51] However, content restrictions for children should not apply equally to adults. Applying child-protective settings indiscriminately could result in censorship of lawful material and unduly restrict adults' freedom to seek, receive and impart information and ideas under art 19(2) ICCPR.

53. This is not an endorsement of age assurance in all its forms. Any age assurance implemented must independently comply with art 17 ICCPR and should:

- use the least privacy-intrusive effective method
- minimise the collection and retention of personal information
- prevent its unrelated use and
- provide accessible means of correcting errors.

54. The appropriate response is therefore to minimise the privacy impacts of differentiated protections, rather than avoid those impacts by extending restrictive child settings to everyone.

Improvements for researchers

55. We support strengthening access to information for researchers. However, the framework could be better tailored to reflect the reality of online safety research. Currently the framework allows for the establishment of data access

schemes for 'approved researchers'.[52] Approved researchers are defined as 'persons employed by an Australian university'.[53] This is in addition to approval by the Minister.[54]

56. Much of the research and advocacy that has gone into supporting online safety regime since the OS Act was first introduced in 2015,[55] has been done by civil society and other advocacy groups.[56] The improvements to researcher access should not be limited to just those employed by Australian universities. Flinders University has rightly described the limitation as 'a highly restrictive view of researchers'.[57] This limited approach consolidates research and data within specific institutions and limits the voices that have been calling for online safety reform.

57. The Australian Government should remove s 205B(1)(a) so that eligibility for approval to undertake online safety research is not limited to employees of Australian universities and can extend to independent researchers, civil society organisations and other research bodies.

58. While broader access to online safety research is important, it should be accompanied by clear safeguards. For example, research conducted through assumed identities could affect users' right to privacy. Clear limits are needed to distinguish observation from manipulation.

59. The Bill should apply safeguards equivalent to those proposed for eSafety's use of sock puppet accounts to approved researchers - limiting their activities to observation, assessment and testing rather than manipulating the digital environment or influencing other users. Engagement with users should be permitted only to the extent necessary to prevent an account from being closed or excluded from the space being studied. A proportionate transparency regime should also require periodic public reporting on the purposes, methods and scope of this research, subject to safeguards for privacy and ongoing research.

Sock puppets

60. The Bill's proposed authorisation for eSafety and approved researchers to use 'sock puppet' requires additional consideration.[58]

61. Allowing the controlled use of sock puppet identities would enable approved researchers and eSafety to test how services operate in practice. It would support more reliable evidence about systemic harms and make it harder for providers to present a selectively compliant version of a service during formal scrutiny.[59]

62. The use of sock puppets is subject to important restrictions. For example, s 205K expressly limits eSafety's engagement with users to what is necessary to prevent account closure. The Explanatory Memorandum explains that the intended role is monitoring, assessment and testing - not participation in online discourse or

ongoing engagement with regular users.[60] eSafety must also maintain policies developed in consultation with the Australian Federal Police and publish an annual statement outlining intended priorities and strategies and summarising the previous year's activities.[61]

63. However, while the Explanatory Memorandum states that the intended role of sock puppet accounts is limited to monitoring, assessment and testing - the legislative provisions themselves appear capable of supporting a broader range of activities. We are concerned that, absent clearer statutory limits, the framework could be interpreted as permitting the use of sock puppet accounts for purposes extending beyond observation and compliance testing. This could include activities that influence online discussions, alter the information environment experienced by users, or promote particular messages or content. While such activities may be pursued with legitimate online safety objectives, the use of undisclosed government-operated accounts to shape online discourse raises distinct concerns relating to freedom of expression, transparency and public trust. We consider that the legislation should more clearly confine the use of sock puppet accounts to investigating, observing and testing compliance with the DDC, and should expressly prohibit their use for the purpose of influencing or participating in online debate or information-sharing activities.

Scope of captured services

64. The DDC casts a broad net to capture the full technology-stack of digital service in an increasingly modern world.[62] The scope of 'online services' departs from the Rickard Review's proposed model, which paired broad coverage with tiered obligations based on a service's reach or risk. For example, the Review proposed that services used by more than 10% of the Australian population would automatically be subject to the highest tier of obligations.[63]

65. The Bill instead applies the DDC broadly to specified categories of online services. As noted by Gilbert + Tobin, even a:

> … simple information website with no user-generated content and no connection to the harms would be required to conduct annual risk assessments, retain records for six years and produce them to the eSafety Commissioner on request.[64]

66. This places a burden on all online services, including those with small users bases or low risk, as they must conduct risk assessments. However, risk assessment and mitigation measures will reflect the level of risk (e.g. smaller, low risk services will have less they need to do on mitigation).

67. While ministerial exemptions for services that pose little risk or are used minimally in Australia can be granted to protect such services – this forms a

piecemeal approach.[65] Further consideration could be given to applying obligations to on online services, proportional to their size.

Generative AI

68. It is reassuring that AI content generation tools are captured, especially in a more broadened form than was set out in the Exposure Draft.[66]

69. This will have a positive impact on the creation and dissemination of synthetic media,[67] including deepfakes. As found in the Commission's upcoming Breaking Trust report, synthetic media can undermine the human rights necessary for the healthy functioning of Australian democracy.

Legislating Codes and Standards and the Basic Online Safety Expectations

70. The Bill will replace the BOSE and sector-specific Industry Codes and Standards framework with an overarching DDC applying across all online services. As found by the Rickard Review, while the BOSE and Industry Codes and Standards have delivered important protections – they are no longer sufficient.[68]

71. The repeal of the Industry Codes and Standards and BOSE and introduction of the overarching duty replace a complex matrix of obligations and regulatory responsibilities.

72. This simplifies compliance for captured businesses and regulatory administration for eSafety. It may also help address inefficiencies under the existing regime.

73. There is clearly a need for simplification given the difficulty in enforcement of the current regime. However, with simplification also comes generalisation across a variety of online services that are very different to one another. What a relevant harm looks like on a search engine could be very different to harms on a service that produces AI generated material (which is why eSafety's guidance discussed below will be so important).

Timing

74. The Bill appears to avoid a formal regulatory gap by providing for the DDC to commence at the same time as the repeal of the BOSE and Industry Codes and Standards. Yet the transition still raises practical concerns. The BOSE and Industry Codes and Standards will remain in force during the 12-month transition period,[69] but the Bill should still ensure that the new framework is operationally ready when those protections are repealed.

75. In particular, eSafety should be required to publish guidance before commencement on risk assessments, proportionate mitigation measures and enforcement priorities. Without clear transitional guidance, there is a risk that

the replacement of the existing framework with a more general duty may produce uncertainty or a temporary weakening of practical protections, even if no formal legal vacuum arises.

eSafety Guidelines

76. Under s 28, the Bill allows eSafety to produce guidelines to assist those responsible for meeting their duty of care. This will assist those across the wide variety of online services to comply with their duty and should encourage approaches to compliance that appropriately balance human rights including safety, privacy, participation and free expression.

Transparency and publication

77. Transparency is fundamental to public trust in the online world, so measures to empower eSafety to demand transparency reports is welcomed.[70] The new transparency regime appears set to replace, and expand the scope of, the existing BOSE reporting framework. Specifically being able to collect information about compliance with the OS Act and other matters which relate to the 'safety of the online service' will better provide confidence to people in Australia that platforms are meeting their obligations without exceeding their mandate. In turn, this transparency would help identify any unintended effects of those obligations on freedom of expression, including where services take measures that exceed what the legislation requires.

Complaints and dispute processes

78. We support s 27G, which requires complaint processes that are available to all people in Australia.[71] Accessible and effective processes will provide users with a direct means of raising concerns about a service, and could assist providers to identify recurring or systemic risks. We also support the proposed power for eSafety to determine, by legislative instrument, the classes of online services to which the obligation applies and the requirements those processes must satisfy.[72]

79. However requirements made under s 27G should expressly address accessibility. Complaint pathways and information about how complaints will be considered should be easy to locate and use, available in accessible formats and written in plain English.

80. It will also be necessary to create child-appropriate pathways for complaint and dispute processes with child-appropriate language to support children's capacity to understand and participate.[73] This will be especially important for online services to ensure they are acting in the bests interests of the child as required under s 3.

81. Participation should reflect each child's evolving capacities, with support and safeguards tailored to their maturity, circumstances and the nature of the complaint.[74] This may include assistance from a parent, trusted adult or independent representative, provided that the arrangements respect the child's views and address any conflicts of interest.[75] The objective is meaningful and supported participation, not an assumption that every child can independently conduct formal proceedings. Equally, age or lack of full legal capacity should not automatically prevent a child from raising a concern.[76]

82. The Explanatory Memorandum should expressly clarify that complaint and dispute resolution pathways under s 27G must be accessible to children, including through child-appropriate language and processes.

83. The complaint and dispute resolution obligations are welcome but may need additional support. The Bill could be strengthened to include an external dispute resolution scheme, so individuals have an accessible, independent body to help them navigate disputes. Without an online services ombudsman, we are concerned that social media platforms may rely too heavily on automated account bans to deal with online harms - with no realistic way of appealing. This will impact people's ability to communicate, work and connect with their communities.

84. Expanding the jurisdiction of the Telecommunications Industry Ombudsman so they also have powers to resolve online services complaints will mean people have timely access to a proven, trusted scheme.

Link deletion and faster removal

85. We support efforts to streamline complaints and removal processes as this is necessary to address viral online harms. However, reducing procedural steps and shortening decision-making timeframes can increase the risk of mistakes and possible overreach – as was seen in *eSafety Commissioner v Baumgarten*.[77]

86. As decision-making processes become faster and more flexible, the risks associated with error, overreach or departures from legislative intent may increase.

Link deletion

87. Link deletion notices are available in relation to specific material already subject to a removal notice: cyber-bullying material targeted at a child, cyber-abuse material targeted at an adult, intimate images, and class 1 material. They do not apply to the full range of material and conduct captured by the DDC.

88. We welcome the expansion of link deletion powers and the requirement to comply within 24 hours, or within a longer period agreed by eSafety.[78] Rapid de-

indexing can be necessary to limit the continuing dissemination of harmful cyber-bullying material or intimate image abuse material.

Design features definition

89. Section 25G(2) allows the Minister to designate additional design features as having negative behavioural impacts. Although this provides flexibility as technology changes, the policy settings should be in the primary legislation (see below at [116]-[118]).

2 Improvements needed

This chapter proposes narrower definitions of harm to children. It also calls for key user empowerment tools in primary legislation, human rights-based risk assessments, political neutrality, and human rights impact assessments for AI-informed moderation tools.

90. We support the principle of the DDC, but the details matter. Its protection of human rights will depend on how clearly its obligations are defined and how services implement them.

Harmful to children definition

91. We are concerned about aspects of the current definition of material and conduct that is 'harmful to children'.[79] While the objective of protecting children from online harm is important, several categories are framed in broad terms which may negatively impact human rights.

Glorifies crime

92. Section 25E(1)(d) places an obligation on online services to ensure that children are protected from material and conduct that 'glorifies crime'. This could extend to material supporting peaceful protest where an aspect of the protest may constitute an offence under Commonwealth, state or territory law.

93. Given the breadth and increasing severity of protest offences across Australia,[80] the provision risks encouraging services to restrict:

- footage of protests
- calls to attend demonstrations or
- discussion of civil disobedience.

94. This creates serious human rights concerns. Risk-averse moderation under s 25E(1)(d) may burden the rights of peaceful assembly and freedom of association under arts 21 and 22 ICCPR. The rights to protest apply to all people without discrimination, including children and young people.[81]

95. This risk is especially significant for young people, who play an important role in protests and other forms of democratic participation. Because they do not have the right to vote, these activities are among the limited ways in which they can express their views and influence public debate.

96. Social media has played a central role in enabling people to organise on issues of public importance.[82] It allows organisers to mobilise people in an effective manner, while also serving as a space for marginalised groups to build community and advocacy where they cannot do so physically (which is especially

important for those in remote or regional communities, people with disability etc).[83] As noted by the Special Rapporteur on the Rights to Freedom of Peaceful Assembly and of Association, technology brings remarkable opportunities to allow rights to both be exercised offline and as a space for individuals to find identity and community - which builds civics movements.[84]

97. The School Strike 4 Climate protests relied heavily on social media platforms to build momentum across jurisdictions and organise protests.[85] Social media has also lowered barriers to participation and allowed young people to build networks that extend well beyond their immediate geographic surroundings. For example, many other youth movements across the globe have been supported by social media including the #RoadSafetyMovement in Bangladesh and the #FeesMustFall campaign in South Africa.[86] Social media has become an integral part of exercising the right to protest for children and young people.

98. The breadth of the provision also creates a risk that restrictions adopted in one jurisdiction will influence content moderation nationally. Although s 25E(1)(d) would not automatically require the removal of every piece of content that breaches a state or territory law, a nationally operating service may respond conservatively to uncertain or differing criminal laws by applying the most restrictive jurisdictional standard across Australia. This could produce unjustified restrictions on lawful political expression in other jurisdictions.

99. As such any limitation on the rights to protest must be prescribed by law and demonstrably necessary and proportionate to a legitimate objective. A framework that encourages the suppression of lawful political communication merely because it is connected with conduct criminalised in one jurisdiction risks exceeding those limits.

100. The exclusions in s 25F provide important safeguards, particularly for material that is in the public interest, but may not fully address these risks. Protest footage, calls to attend demonstrations and discussion of civil disobedience could fall within those exclusions, yet their protection would depend on how services assess the purpose and context of each piece of content. Faced with uncertainty about whether material 'glorifies crime', risk-averse services may restrict content rather than assess whether an exclusion applies. The exclusions do not remove the need to narrow the underlying provision so that support for peaceful protest is not treated as harmful merely because the conduct involved may constitute an offence.

Encourages seriously harmful practices

101. Protecting children from material encouraging seriously harmful practices pursues a legitimate aim under art 19(3) ICCPR and gives effect to the obligation under art 17(e) of the *Convention on the Rights of the Child* to develop safeguards against information injurious to their wellbeing. However, children also hold

rights to expression and access to information – and General Comment No. 25 requires online content restrictions to be necessary, proportionate and consistent with their evolving capacities.[87] We are concerned that 'seriously harmful practices' is too broad and may impose a disproportionate impact on freedom of expression.

102. Unlike s 25D(1)(i), this provision does not require seriously harmful practices to be 'reasonably likely' to cause harm. The Explanatory Memorandum's reference to injury and death explains its protective purpose, but does not establish a causal connection to a likelihood to cause harm.[88] This creates a concern under General Comment No. 34 as online services need an objective basis for distinguishing encouragement of dangerous conduct from discussion or advocacy concerning a disputed practice.[89]

103. A preventative restriction should be connected to a practice with a reasonable likelihood of serious harm.[90] While s 25F is an important safeguard – it does not cure uncertainty in the underlying definition.

104. Additionally we are concerned by the idea of Big Tech platforms being the arbiters of what constitutes 'seriously harmful practices' for children. The addition of 'seriously' in this version of the DDC narrows s 25E(1)(d), but does not resolve uncertainty.[91]

105. As noted in their submission on the Exposure Draft in relation to the definition of 'harmful to children' Digital Rights Watch stated:

> Further there is a clear risk that hate groups in society will attempt to use this legislation to de-platform those that they hate, by claiming that there are potential harms to children. Already, platforms frequently deplatform LGBTI content, or sexual health and education messaging, as they perceive it easier to do so than to contest its validity.[92]

106. We are particularly concerned about how anti-trans narratives about the harm to children could be weaponised by the leadership of certain platforms to restrict important information and advocacy.

Recommendation: The Australian Government should remove the 'encourages seriously harmful practices' limb of s 25E(1)(d) or clearly define its scope in primary legislation to reduce the risk of unjustified restrictions on lawful expression and access to information.

May cause serious harm to children

107. Clarification in the Explanatory Memorandum that 'serious harm' in the Bill refers to the definitions contained in s 5 OSA is welcomed.[93] However, we remain concerned about the low threshold adopted in s 25E(1)(f).

108. The principles governing restrictions on freedom of expression are discussed at paragraphs [22]-[30] above. We consider 'may cause' serious harm in s 25E(1)(f) too broad. It does not clearly explain which practices are covered or how likely they must be to cause serious harm. This leaves platforms to make those judgments, creating a risk that they will restrict lawful expression based on unclear understandings of harm.

109. We support preventing harm to children, but restrictions must be clearly defined and no broader than necessary. The public-interest exclusions in s 25F help protect lawful expression, but does not resolve the uncertainty in this provision. The Bill should expressly limit this limb to practices reasonably likely to cause serious harm to children.

Recommendation: The Australian Government should amend s 25E(1)(f) to replace 'may cause' with 'is reasonably likely to cause', to provide a clearer threshold and reduce the risk of overreach in content moderation.

Hostile attitudes towards women or gender equality

110. We understand 'gender equality' to encompass equality for people of all genders, including women, men (whether cisgender or transgender) and non-binary people. This is consistent with Office of the United Nations High Commissioner for Human Rights' approach to gender equality, which expressly includes people of diverse gender identities and seeks equal participation and more equal power relations for all genders.[94]

111. Misogyny is increasingly recognised as a factor that can contribute to harm to women at a time when Australia appears particularly vulnerable to action by those who have been radicalised online. Research by the University of Melbourne found that hostility towards women was a stronger predictor of extremist attitudes than political ideology, religion or ethnicity – identifying misogyny as a factor that may cut across multiple forms of violent extremism.[95] The point here is that hostile attitudes aren't the harm itself – but are leading people online towards increasingly dangerous actions both offline and online. While the DDC is an important step to address drivers of these harms, it is not a panacea – focus must still remain on education and engagement with young people.

112. The principles governing restrictions on freedom of expression are discussed at paragraphs [22]-[30] above. Protecting children from material that encourages gender-based discrimination is a legitimate objective, consistent with CEDAW's requirement to address prejudices and stereotypes based on the inferiority or superiority of either sex.[96] However, 'hostile attitudes' in s 25E(1)(c) is unclear and risks capturing criticism of particular gender equality policies, rather than material that encourages discrimination or harm. People are free to hold

opinions, even hostile ones. The Bill regulates how those views are expressed, but any restriction must still meet the requirements of art 19(3) ICCPR.

113. We support the protective purpose of s 25E(1)(c), but consider that 'hostile attitudes' needs a clearer statutory boundary. Unlike s 25D(1)(i), this provision does not require serious harm to be reasonably likely. Children may need broader protection, but the Bill must still make clear what harm justifies restricting their access to information.

114. The Bill should clarify the connection required between the attitudes promoted and harm to children, while protecting discussion of gender. The exclusions in s 25F assist, but cannot replace a clear underlying definition or the requirement to use the least restrictive effective response.

Removal of Minister's delegated powers

115. The removal of the Minister's power to prescribe additional categories of material and conduct that are harmful to children is a positive amendment.

User empowerment tools

116. The user empowerment tools should be set out in the primary legislation to begin with – especially as Government has promoted the DDC on the basis of giving users more control over their digital experience.[97] Our view is that the policy settings for the DDC should be set out clearly in the primary legislation.

117. For example, the *Online Safety (User Empowerment) Determination 2026* (Cth) is currently out for consultation. However, this appears to have only been offered to organisations with past engagement on the DDC. The ability to provide feedback on this instrument should have been offered to all people, regardless of past engagement.

118. Consistent with the Law Council of Australia, we believe that given the importance of user empowerment tools this should be contained in the primary legislation and subject to parliamentary scrutiny.[98] Should this not happen, there should be an agreed process of public stakeholder consultation and adequate opportunities to publicly scrutinise any proposal prior to the introduction.[99]

Recommendation: The Australian Government should establish key user empowerment tools in primary legislation, subject to parliamentary scrutiny. If these tools are instead prescribed through ministerial rules, the Government should ensure public stakeholder consultation and adequate opportunities for public scrutiny of proposed rules before they are made.

Risk assessments

119. We support the requirement for online services to conduct risk assessments and implement mitigation strategies to address risks of harm posed by the service in Australia. However, the framework does not expressly incorporate human rights considerations. As a result, services are not required to identify or balance the impact of mitigation measures on rights such as equality and non-discrimination, privacy, or freedom of expression.

120. The Bill should be amended to expressly require that risk assessments must address cumulative and contextual factors that may contribute to harm, including the interaction between content, amplification mechanisms and platform design features. Human rights-based criteria or principles should guide proportionate risk mitigation, including respect for privacy, safety and freedom of expression (which includes the right to seek and receive information).

121. In conducting risk assessments, we also recommend that captured services conduct specific child rights impact assessments. This would not only allow for contextual analysis of risk but would also ensure compliance with online services overriding obligation to act in the best interests of children – as provided for in s 3.

Political neutrality

122. We are concerned that public debate about the DDC has at times linked the proposed reforms to particular political parties and the reach of their messages online.[100] The purpose of the DDC must be to reduce foreseeable harm arising from the design and operation of online services, not to influence which political views people encounter.

Moderation and HRIA

123. Online services are likely to deploy or expand AI-informed content moderation tools to meet their obligations under the DDC. These tools must not improve safety at the expense of lawful expression, interference with privacy or discriminatory outcomes. The Bill should require services to conduct a human rights impact assessment (HRIA) before deploying these tools or making significant changes to them to limit human rights violations.

124. A HRIA is a systematic process used to identify, evaluate and address the potential effects of a policy, law or systems on people's human rights.[101] It ensures that new measures (such as content filtering) do not inadvertently harm or discriminate against individuals or groups, and that any risks are properly managed and mitigated.[102]

3 One-year statutory review

The Bill should require a statutory review one year after the DDC commences. The review should examine whether the harm categories should expressly include hate speech and whether people harmed by a breach should have a private right of action.

125. Given the significance of the reform, the Bill should require a statutory review one year after the duty commences. The review should consider, in particular, the operation of the act, whether the categories of harm in proposed ss 25D and 25E should expressly include hate speech, if the content exclusions in s 25F are appropriately calibrated and whether people who suffer harm because an online service breaches the duty should have a private right of action. We would welcome the opportunity to provide expertise to this review.

126. We acknowledge that these are complex issues, involving practical implementation considerations. There will be different views on these issues (including within the Commission) such as on the balancing of human rights considerations and on the suite of regulatory powers that should be available to support compliance with the DDC.

Recommendation: The Australian Government should amend the Bill to require a statutory review one year after the digital duty of care commences, including consideration of, *inter alia*, whether ss 25D and 25E should expressly include hate speech and whether people harmed by a breach of the duty should have a private right of action.

Endnotes

[1] Senate Legal and Constitutional Affairs References Committee, Parliament of Australia, *Adequacy of Existing Offences in the Commonwealth Criminal Code and of State and Territory Criminal Laws to Capture Cyberbullying* (Report, March 2018) rec 8, viii [5.28] & 62-63 [5.23]-[5.28].

[2] House of Representatives Select Committee on Social Media and Online Safety, Parliament of Australia, *Social Media and Online Safety* (Report, March 2022) rec 20 xxii & 182-184 [5.82]-[5.88].

[3] Joint Select Committee on Social Media and Australian Society, Parliament of Australia, *Social Media: The Good, the Bad, and the Ugly* (Final Report, November 2024) rec 2, xi [5.141] & 118 [5.129].

[4] Senate Environment and Communications Legislation Committee, Parliament of Australia, *Online Safety Amendment (Social Media Minimum Age) Bill 2024 [Provisions]* (Report, November 2024) rec 1, ix [2.105] & 28-29 [2.96]-[2.105].

[5] Senate Environment and Communications References Committee, Parliament of Australia, *Internet Search Engine Services Online Safety Code and under 16 Social Media Ban* (Report, November 2025) rec 2, xi [3.48] & 42-44 [3.8]-[3.14].

[6] Lynelle Briggs, *Report of the Statutory Review of the Enhancing Online Safety Act 2015 and the Review of Schedules 5 and 7 to the Broadcasting Services Act 1992 (Online Content Scheme)* (Report, October 2018) 42 rec 2(a).

[7] Delia Rickard, *Report of the Statutory Review of the Online Safety Act 2021* (Report, October 2024) rec 4, 21 & 46-63.

[8] See Australian Human Rights Commission ('AHRC'), *Statutory Review of the Online Safety Act 2021* (Submission to the Department of Infrastructure, Transport, Regional Development, Communications and the Arts, 28 June 2024) rec 14, 17-18 [65]-[67].

[9] AHRC, *Online Safety Amendment (Social Media Minimum Age) Bill 2024* (Submission to the Senate Standing Committees on Environment and Communications, 22 November 2024) 8; AHRC, 'Explainer: Social Media Ban for Under-16s in Australia' (Webpage) <https://humanrights.gov.au/about-us/news/proposed-social-media-ban-under-16s-australia>.

[10] AHRC, *Technology-Facilitated Gender-Based Violence* (Submission to the Human Rights Council Advisory Committee, 28 March 2025) 7 [21]-[22].

[11] AHRC, *Search Engine Safety Code and Social Media Ban* (Submission to the Senate Environment and Communications References Committee, October 2025) 25-26 [87]-[91].

[12] See generally AHRC, *A Digital Duty of Care for Australia* (Submission, 22 July 2026); See also AHRC, *Digital Duty of Care Exposure Draft Bill* (Submission to the Department of Infrastructure, Transport, Regional Development, Communications, Sport and the Arts, 23 September 2026).

[13] *Online Safety Amendment (Digital Duty of Care) Bill 2026* (Cth) ('Draft Bill') ss 25C & 26(1).

[14] Draft Bill s 25C(1)(a)-(c).

[15] See *Online Safety Amendment (Digital Duty of Care) Bill 2026* (Cth) Exposure Draft Bill ('Exposure Draft Bill') s 25H(e)(ii).

[16] AHRC, *Digital Duty of Care Exposure Draft Bill* (Submission to the Department of Infrastructure, Transport, Regional Development, Communications, Sport and the Arts, 23 September 2026) 4 [3], 5–6 [9]–[16].

[17] Grattan Institute, *For the People: Future-proofing Australia's democracy* (Report, 2026) 24.

[18] University of Canberra (News and Media Research Centre), *Digital News Report: Australia 2025* (Report, June 2025) 11.

[19] AHRC, *Digital Duty of Care Exposure Draft Bill* (Submission to the Department of Infrastructure, Transport, Regional Development, Communications, Sport and the Arts, 23 September 2026) 5–6 [9]–[16].

8.

[20] Committee on the Elimination of Discrimination against Women, *General Recommendation No. 35 on Gender-Based Violence against Women, Updating General Recommendation No. 19*, UN Doc CEDAW/C/GC/35 (26 July 2017) 9 [24(b)], 13 [30] .

[21] Human Rights Committee, *General Comment No. 34: Article 19—Freedoms of Opinion and Expression*, UN Doc CCPR/C/GC/34 (12 September 2011) paras 6 [22], 6-7 [25], 8 [33]–[35].

[22] Draft Bill s 25D(1)(i); cf Exposure Draft Bill s 25C(1)(f).

[23] Explanatory Memorandum, *Online Safety Amendment (Digital Duty of Care) Bill 2026* (Cth) 59.

[24] Human Rights Committee, *General Comment No. 34: Article 19—Freedoms of Opinion and Expression*, UN Doc CCPR/C/GC/34 (12 September 2011) 3 [11].

[25] ICCPR art 19(3)(b)

[26] Human Rights Committee, *General Comment No. 34 34: Article 19—Freedoms of Opinion and Expression*, UN Doc CCPR/C/GC/34 (12 September 2011), 6 [25].

[27] Frank La Rue, Annual Report of the Special Rapporteur on the Promotion and Protection of the Right to Freedom of Opinion and Expression, UN Doc A/HRC/14/23 (20 April 2010) 13-15 [79]-[81]; See also Human Rights Committee, General Comment No 34 (Article 19: Freedom of opinion and expression) UN Doc CCPR/C/GC/34 (12 September 2011) 6 [22].

[28] ICCPR art 19(3)(a); Human Rights Committee, *General Comment No. 34: Article 19—Freedoms of Opinion and Expression*, UN Doc CCPR/C/GC/34 (12 September 2011) 7 [28]; See also Explanatory Memorandum, *Online Safety Amendment (Digital Duty of Care) Bill 2026* (Cth) 55–56.

[29] Explanatory Memorandum, *Online Safety Amendment (Digital Duty of Care) Bill 2026* (Cth) 16–17. ; See also Delia Rickard, *Report of the Statutory Review of the Online Safety Act 2021* (Report, October 2024) 13.

[30] Noting that Explanatory Memorandum, *Online Safety Amendment (Digital Duty of Care) Bill 2026* (Cth) 72 provides that services should consider cumulative impacts as part of risk assessments.

[31] Human Rights Committee, *General Comment No. 34: Article 19—Freedoms of Opinion and Expression*, UN Doc CCPR/C/GC/34 (12 September 2011) 7 [28] explaining that the 'rights of others' include the rights of 'individual persons or of individual persons as members of a community'; Committee on the Elimination of Racial Discrimination, *General Recommendation No. 35: Combating Racist Hate Speech*, UN Doc CERD/C/GC/35 (26 September 2013) 4 [13(b)–(d)].

[32] Committee on the Elimination of Racial Discrimination, *General Recommendation No. 35: Combating Racist Hate Speech*, UN Doc CERD/C/GC/35 (26 September 2013) 4 [12], 5 [15], 6 [20].

[33] See e.g. Royal Commission on Antisemitism and Social Cohesion, *Transcript of Proceedings*, Sydney, 30 June 2026, 91; Royal Commission on Antisemitism and Social Cohesion, *Transcript of Proceedings*, Sydney, 2 July 2026, 291.

[34] AHRC, *Equal Identities: A Human Rights Review of the Experiences of Trans and Gender Diverse People in Australia* (Report, March 2026) 45–50.

[35] Draft Bill s 25D(1)(m); cf Exposure Draft Bill s 25C(1)(j).

[36] See e.g. AHRC, *A Digital Duty of Care for Australia* (Submission, 22 July 2026) 5 [6]; AHRC, *Digital Duty of Care Exposure Draft Bill* (Submission, 23 September 2026) 7-8 [18]–[20] & [24].

10.

8.

[37] Draft Bill s 3.

[38] See generally *Privacy (Children's Online Privacy) Code 2026* (Exposure Draft, 31 March 2026) ss 10–11.

[39] My Image, My Choice, *The Issue* (webpage) <https://myimagemychoice.org/>.

[40] See generally Deeptrace Labs, *The State of Deepfakes: Landscape, Threats, and Impact* (Report, 2019).

[41] Julie Inman Grant, 'Addressing deepfake image-based abuse', *eSafety* (Online, 24 July 2024) <https://www.esafety.gov.au/newsroom/blogs/addressing-deepfake-image-based-abuse>.

[42] Nadeem Badshah, 'Nearly 4,000 celebrities found to be victims of deepfake pornography', *The Guardian* (online, 22 March 2024) <https://www.theguardian.com/technology/2024/mar/21/celebrities-victims-of-deepfake-pornography>.

[43] *Convention on the Elimination of All Forms of Discrimination against Women* art 7.

[44] Committee on the Elimination of Discrimination against Women, *General Recommendation No. 35 on Gender-Based Violence against Women, Updating General Recommendation No. 19*, UN Doc CEDAW/C/GC/35 (26 July 2017) 7 [20] & 9 [24(b)].

[45] Draft Bill ss 86A-86D.

[46] See e.g. *Privacy (Children's Online Privacy) Code 2026* (Exposure Draft, 31 March 2026); *Online Safety Act 2021* (Cth) pt 4A; *Internet Search Engine Services Online Safety Code (Class 1C and Class 2 Material)* (2025).

[47] Committee on the Rights of the Child, *General Comment No. 25 (2021) on Children's Rights in Relation to the Digital Environment*, UN Doc CRC/C/GC/25 (2 March 2021) 4 [19].

[48] Committee on the Rights of the Child, *General Comment No. 25 (2021) on Children's Rights in Relation to the Digital Environment*, UN Doc CRC/C/GC/25 (2 March 2021) 4 [19]-[20].

[49] Lorraine Finlay & Patrick Hooton, 'Fast Law, Fragile Rights: The Human Rights Consequences of Australia's Rapid Social Media Reforms' (2025) 13(1) *Griffith Journal of Law & Human Dignity* 1, 20-21.

[50] See generally eSafety, *Social Media Minimum Age: Compliance update* (Report, March 2026); Leonardo Bursztyn, et al., *Why Bans Fail: Tipping Points and Australia's Social Media Ban* (Research brief, University of Chicago, 21 May 2026).

[51] Digital Rights Watch, *Online Safety Amendment (Digital Duty of Care) Bill 2026* (Submission to the Department of Infrastructure, Transport, Regional Development, Communications, Sport and the Arts, September 2026) 6.

[52] Draft Bill ss 205B–205C.

[53] Draft Bill s 205B(1)(a).

[54] Draft Bill s 27A(1).

[55] The Act was previously known as the *Enhancing Online Safety Act 2015* (Cth).

[56] See e.g, advocacy conducted by Teach us Consent as part of their Fix Our Feeds campaign.

[57] Flinders University, *Submission on the Online Safety Amendment (Digital Duty of Care) Bill 2026* (Submission, 22 September 2026) 7 [2.15.1]

[60] Explanatory Memorandum, *Online Safety Amendment (Digital Duty of Care) Bill 2026* (Cth) 98–99.

[61] *Online Safety Amendment (Digital Duty of Care) Bill 2026* (Cth) sch 2 item 43, inserting ss 205K(3), 205M(1)–(2); Explanatory Memorandum, *Online Safety Amendment (Digital Duty of Care) Bill 2026* (Cth) 99–100.

[62] Draft Bill s 25B.

10.

8.

[63] Delia Rickard, *Report of the Statutory Review of the Online Safety Act 2021* (Report, October 2024) rec 7, 21.

[64] Gilbert + Tobin, 'Australia's Digital Duty of Care Takes Shape — From Algorithms to Everything Else' (WebPage) < https://www.gtlaw.com.au/insights/australias-digital-duty-of-care-takes-shape-from-algorithms-to-everything-else>.

[65] Draft Bill s 25B(3).

[66] Draft Bill s 25B(1)(h); cf Exposure Draft Bill s 25A(1)(j).

[67] An umbrella term for any AI-generated or altered photo, video or audio (including deepfakes).

[68] Delia Rickard, *Report of the Statutory Review of the Online Safety Act 2021* (Report, October 2024) 49-50 & 63.

[69] Draft Bill s 2(1) table item 3, sch 2 item 13, sch 3 item 5; Explanatory Memorandum, Online Safety Amendment (Digital Duty of Care) Bill 2026 (Cth) 2, 44–45.

[70] Draft Bill s 192B.

[71] Cf Exposure Draft Bill s 26F(1)(a).

[72] Draft Bill s 27G.

[73] See e.g. Australian Research Council Centre of Excellence for the Digital Child, *Response to the Online Safety Amendment (Digital Duty of Care) Bill 2026* (Submission, September 2026) 3–4.

[74] Committee on the Rights of the Child, *General Comment No. 25*, UN Doc CRC/C/GC/25 (2 March 2021) 4 [19]–[21]; *General Comment No. 12*, UN Doc CRC/C/GC/12 (20 July 2009) 11 [29]–[30].

[75] Committee on the Rights of the Child, *General Comment No. 12*, UN Doc CRC/C/GC/12 (20 July 2009) 12 [35]–[37].

[76] Committee on the Rights of the Child, *General Comment No. 12*, UN Doc CRC/C/GC/12 (20 July 2009) 12 [35]–[39];

[77] (2026) 315 FCR 44.

[78] *Online Safety Amendment (Digital Duty of Care) Bill 2026* (Cth) sch 1 items 27–29, inserting ss 72A(1)(d), 72B, 82(1)(d), 82A, 92A(1)(d) & 92B.

[79] Draft Bill s 25E.

[80] See generally David Mejia-Canales, *Protest in Peril: Our Shrinking Democracy* (Human Rights Law Centre, 2 June 2024).

[81] Report of the Special Rapporteur on the Rights to Freedom of Peaceful Assembly and of Association, UN Doc A/HRC/62/45 (21 April 2026) 7 [31]; *Convention on the Rights of the Child* art 15; *International Covenant on Civil and Political Rights* arts 21–22.

[82] David Kaye, *Surveillance and human rights* (Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression, 2019, UN Doc A/HRC/41/35) 4 [11]; see also Maina Kiai, *Report of the Special Rapporteur on the rights to freedom of peaceful assembly and of association* (2012, UN Doc A/HRC/20/27); Clément Nyaletsossi Voule, *Report of the Special Rapporteur on the rights to freedom of peaceful assembly and of association* (2018, UN Doc A/HRC/38/34).

[83] David Kaye, *Surveillance and human rights* (Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression, 2019, UN Doc A/HRC/41/35) 4 [11]; see also Maina Kiai, *Report of the Special Rapporteur on the rights to freedom of peaceful assembly and of association* (2012, UN Doc A/HRC/20/27); Clément Nyaletsossi Voule, *Report of the Special Rapporteur on the rights to freedom of peaceful assembly and of association* (2018, UN Doc A/HRC/38/34).

10.

8.

[84] David Kaye, *Surveillance and human rights* (Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression, 2019, UN Doc A/HRC/41/35) 6 [21].

[85] See generally Shelley Boulianne, Mireille Lalancette and David Ilkiw, '"School Strike 4 Climate": Social Media and the International Youth Protest on Climate Change' (2020) 8(2) Media and Communication 208.

[86] David Kaye, *Surveillance and human rights* (Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression, 2019, UN Doc A/HRC/41/35) 6 [22].

[87] CRC arts 5, 13, 17; Committee on the Rights of the Child, *General Comment No. 25 (2021) on Children's Rights in Relation to the Digital Environment*, UN Doc CRC/C/GC/25 (2 March 2021) 4 [19]–[21], 9-10 [54]–[60].

[88] Explanatory Memorandum, *Online Safety Amendment (Digital Duty of Care) Bill 2026* (Cth) 55.

[89] Human Rights Committee, *General Comment No. 34: Article 19—Freedoms of Opinion and Expression*, UN Doc CCPR/C/GC/34 (12 September 2011) 6-7 [25].

[90] See generally Human Rights Committee, *General Comment No. 34: Article 19—Freedoms of Opinion and Expression*, UN Doc CCPR/C/GC/34 (12 September 2011) 8 [33]–[35]; Explanatory Memorandum, *Online Safety Amendment (Digital Duty of Care) Bill 2026* (Cth) 16–17.

[91] Explanatory Memorandum, *Online Safety Amendment (Digital Duty of Care) Bill 2026* (Cth) 58–59.

[92] Digital Rights Watch, *Online Safety Amendment (Digital Duty of Care) Bill 2026* (Submission to the Department of Infrastructure, Transport, Regional Development, Communications, Sport and the Arts, September 2026) 11.

[93] Explanatory Memorandum, *Online Safety Amendment (Digital Duty of Care) Bill 2026* (Cth) 60.

[94] Office of the United Nations High Commissioner for Human Rights, *OHCHR and Women's Human Rights and Gender Equality* (Webpage) <https://www.ohchr.org/en/women>.

[95] Sara Meger, *Misogyny, Online Radicalisation, and Youth Extremism in Australia* (Research Report, The University of Melbourne, June 2026) 23.

[96] CEDAW art 5(a); See also Committee on the Elimination of Discrimination against Women, *General Recommendation No. 35 on Gender-Based Violence against Women, Updating General Recommendation No. 19*, UN Doc CEDAW/C/GC/35 (26 July 2017) 13 [30(b)].

[97] See e.g. Anthony Albanese and Anika Wells, 'My Feed, My Way' (Joint Media Release, 8 September 2026).

[98] Law Council of Australia, *Online Safety Amendment (Digital Duty of Care) Bill 2026: Exposure Draft* (Submission to the Department of Infrastructure, Transport, Regional Development, Communications, Sport and the Arts, 23 September 2026) 12.

[99] Law Council of Australia, *Online Safety Amendment (Digital Duty of Care) Bill 2026: Exposure Draft* (Submission to the Department of Infrastructure, Transport, Regional Development, Communications, Sport and the Arts, 23 September 2026) 12.

[100] See e.g. Nick Newling, 'Algorithm shake-up will hurt right-wing parties, censorship fight heats up', *Sydney Morning Herald* (online, 9 September 2026), <https://www.smh.com.au/politics/federal/algorithm-shake-up-will-hurt-right-wing-parties-censorship-fight-heats-up-20260908-p60vd0.html>; Laurence Karacsony

&

Oscar Godsell , 'Greens Senator David Shoebridge claims Digital Duty of Care is needed to stop 'amplifying' One Nation, Coalition rhetoric', *News24* (online, 17 September 2026) <https://www.news24.com.au/politics/australian-

10.

8.

politics/greens-senator-david-shoebridge-claims-digital-duty-of-care-is-needed-to-stop-amplifying-one-nation-coalition-rhetoric/news-story/b7127733162a422542e97eadd712f8b2>.

[101] AHRC & National Australia Bank, *Human Rights Impact Assessment Tool: AI-informed Decision-making Systems in Banking* (September 2023) 7-8.

[102] See e.g. AHRC, *Technology and Human Rights* (Final Report, 2021) 55-59 in the context of artificial intelligence.