ORYGEN SUBMISSION

THE ONLINE SAFETY AMENDMENT (DIGITAL DUTY OF CARE) BILL 2026

Orygen welcomes the opportunity to provide a submission to the Online Safety Amendment (Digital Duty of Care) Bill 2026. Orygen's submission will address issues in relation to online safety and the mental health and wellbeing of young people.

Orygen supports the Online Safety Amendment (referred to throughout this submission as the Digital Duty of Care) as an important regulatory measure to reduce young people's exposure to online harms, strengthen the safety of online environments and complement broader government investments in youth mental health and wellbeing.

ABOUT ORYGEN

Orygen is the world's leading research and knowledge translation organisation focusing on mental ill-health in young people. At Orygen, our leadership and staff work to deliver cutting-edge research, policy development, innovative clinical services, and evidence-based training and education to ensure that there is continuous improvement in the treatments and care provided to young people experiencing mental ill-health.

Orygen has undertaken significant work to make social media platforms safer spaces for young people. For example, our world-leading suicide prevention team developed the #chatsafe program, which teaches young people, educators and parents about safe online conversations about suicide and self-harm. These guidelines were developed for young Australians independently of any social media platforms. We have developed further regional adaptations in partnership with global health organisations since then, bringing our youth guidelines to 18 countries, and our parents and carers resource to 27 countries in total. These world-first guidelines have now been downloaded over 1 million times, and our social media campaigns have reached 12.6 million people.

CONTACT DETAILS

For further information, please contact:

Vivienne Browne
Director, Policy and Engagement
Vivienne.browne@orygen.org.au

OVERVIEW

With rates of self-harm and suicidal ideation among young people increasing,(7) there is a pressing need to strengthen online safety for young people. Social media companies should implement stronger platform safeguards, including partnering with mental health organisations and employing mental health-qualified content moderators. They should also ensure that algorithms are designed to

The #chatsafe® project is funded by the Australian Government under the National Suicide Prevention Leadership and Support Program. Funding from Meta supported Orygen's adaptation of these guidelines for implementation in 10 countries in 2019 and the development of parent-focused guidelines across 15 countries in 2022.

reduce, rather than amplify, exposure to harmful content.(5) Incentivising these actions will require legislated requirements, regulation and monitoring (including requirements for platforms to report on how they are enacting content policies), as well as clear consequences for breaches, defined and enforced by governments.

Exposure to online harms is not an unavoidable feature of online environments. Decisions about platform design, recommender systems and safety features can significantly influence young people's exposure to harmful content.(1, 2) The Digital Duty of Care represents an important mechanism for requiring online services to take proactive steps to identify, prevent and mitigate foreseeable harms arising from the design and operation of their platforms.

While many platforms have policies prohibiting self-harm and suicide-related content, this content remains readily accessible on major platforms and can be rapidly amplified through recommender systems. The existence of policies alone is insufficient to prevent or reduce harm. Research indicates that current approaches could be improved through broader application, greater transparency and strengthened implementation. In addition, existing measures are mostly platform-specific and few have been empirically tested for acceptability or effectiveness.(3)

The Digital Duty of Care should require online services to demonstrate that measures intended to prevent exposure to harmful content, including self-harm and suicide-related content, are effective in practice. This should include appropriately trained moderation systems, regular evaluation of safety measures, and ongoing collaboration with social media companies, mental health experts, researchers, policymakers and young people with lived experience.(3) This would help ensure that safety measures are evidence-based, responsive to user needs, and effective across different developmental stages.

ONLINE SAFETY OF YOUNG PEOPLE

IMPACTS ON MENTAL HEALTH AND WELLBEING

Orygen supports the Bill's recognition that exposure to online harms can have significant consequences for young people's mental health and wellbeing, and that online services have an important role in preventing and mitigating risks associated with online abuse, discriminatory content, and content that promotes self-harm, eating disorders and suicide.

Social media can offer important benefits for young people's mental health and wellbeing, including opportunities for social connection, help-seeking, and access to mental health information and resources.(2, 4-6) Young people are significantly more likely than adults to seek support through social media, with one in five reporting that they had sought support online, compared with just five per cent of adults.(7) Online platforms can reduce traditional barriers to mental health support.(6, 7) Around one-quarter of young people and almost half (46.8%) of LGBTQIA+ young people report turning to online support because they were unable to access help offline.(8)

Despite these positive aspects, online environments can expose young people to mental health harm, particularly where platform design and engagement-driven algorithms prioritise user engagement over safety.(2) Distressing content often attracts high levels of online engagement, and despite ongoing calls for regulation, recommender systems continue to facilitate its rapid spread, with potential implications for young people's mental health and wellbeing.(2, 6, 7, 9)

A key concern is young people's exposure to unregulated and uncensored content, including graphic material depicting self-harm and suicide. A national survey conducted by Orygen researchers revealed that approximately 75 per cent of young people had been exposed to suicide-related content online, a rate higher than that reported by adults. Exposure was particularly high among LGBTQIA+ young people, with 90.8 per cent reporting exposure to suicide-related content online.(8) This is particularly concerning given the potential mental health impacts of exposure to self-harm and suicide-related content, which has been associated with increased rates of depression, anxiety and suicidal behaviour

among young people.(2) Research found that exposure was higher among young people with lived experience of mental ill-health and those under 16 years of age. Most young people reported encountering this content unintentionally, with many indicating that it worsened their mood and, for a small number, contributed to increased self-harm.(8)

Social media use can also negatively affect young people's body image and is associated with unrealistic body ideals.(6) Recommender algorithms have been documented promoting pro-anorexia content(7) and suggesting anorexia community accounts, including 'AnaCoaches', as 'friends' to children's accounts.(10) Exposure to this content and these communities can normalise disordered eating and prompt young people to imitate harmful behaviours.(10, 11)

These findings highlight why it is important to engage young people directly in the development of online safety responses.(10) Young people are among the most active users of online platforms and are best placed to describe how they encounter harmful content, what helps them stay safe, and where current protections fall short. Their experiences also differ, with exposure higher among those with lived experience of mental ill-health and those under 16 years of age.(5) Because many young people encounter harmful content unintentionally, and many also rely on online spaces to seek support,(5) safety measures designed without their input risk being ineffective, or limiting the benefits that online environments offer. Involving young people in the design and evaluation of online safety measures helps ensure that responses reflect their lived experience, are credible and accessible to them, and are more likely to reduce harm without undermining help-seeking and social connection.

TECHNOLOGY-FACILITATED ABUSE AND/OR HARM

Young people are increasingly affected by emerging forms of technology-facilitated abuse, including non-consensual image sharing, AI-generated abusive or sexualised content and online harassment and abuse.(12-14) Experiences of technology-facilitated abuse are common among young people. Seven in ten young people (72%) aged 18–24 have experienced technology-facilitated abuse in their lifetime, the highest proportion of any age group.(15) Technology-facilitated abuse presents a risk for young people's wellbeing, with two-thirds (67%) reporting negative impacts on their mental health.(16)

Orygen supports amendments to strengthen the child cyberbullying and adult cyber abuse schemes by reducing the timeframe for the eSafety Commissioner to issue removal notices following a complaint, and for services to comply with those notices, from 48 hours to 24 hours. These reforms recognise that delays in addressing harmful content can exacerbate psychological distress and increase the risk of harm associated with technology-facilitated abuse, particularly where young people are exposed to suicide-related harassment.(14, 17)

The Digital Duty of Care should require online services to take proactive action to prevent and mitigate foreseeable harms associated with technology-facilitated abuse, cyberbullying and image-based abuse and harm before they occur, rather than relying primarily on complaints-based responses after harm has occurred. While complaints-based responses are an important safeguard, they may occur too late to prevent significant impacts on mental health and wellbeing. This includes limiting features that encourage compulsive and prolonged use and a focus on physical appearance.(18)

THE USE OF AI ASSISTANTS AND COMPANIONS

The use of AI assistants and companions have become a common feature of young people's online lives.(19) Around 4 in 5 young people aged 10–17 have used one or more AI assistants, with ChatGPT, Gemini and Copilot among the most commonly used. Although these technologies have the potential to benefit young people, they also present risks to wellbeing and safety, including the promotion of harmful behaviours, privacy risks, inaccurate health information and exposure to violent or sexually explicit material.(19, 20)

Young people are increasingly turning to AI assistants and companions for emotional support and mental health information. Among young people aged 10–17 years, 22 per cent reported using these

tools to discuss feelings or challenges, 20 per cent sought advice about their mental health and wellbeing, and 29 per cent used them for personal support or connection. These interactions have implications for online safety. AI assistants and companions often do not respond appropriately to mental health-related disclosures and, in some cases, may provide responses that reinforce or encourage self-harm or suicidal thoughts among young people.(19) Existing safeguards also seem to be insufficient, with reporting mechanisms often not visible or accessible to young people.(20) eSafety research found that 1 in 5 (20%) young people reported having experienced potentially inappropriate or harmful interactions with these tools,(20) highlighting the importance of ensuring AI systems implement safeguards to assess and manage risks to young people's mental health and wellbeing, including harms arising from sustained interactions, such as manipulation, dependency and the transfer of sensitive personal information.(21)

Recommendations

The Digital Duty of Care should establish clear obligations for online service providers to prevent and address young people's exposure to self-harm, suicide-related and other harmful content and proactively assess and mitigate the risk for technology facilitated abuse.

These obligations should apply consistently across digital services, including social media platforms and emerging generative-AI technologies.(3)

Safety by design principles should underpin all online environments accessed by young people. Young people must have a meaningful role in this process, including the development and evaluation of online safety measures.

The Digital Duty of Care should include amendments to:

- Provide clear and explicit guidance within the Digital Duty of Care on industry expectations for preventing and mitigating exposure to eating disorder, self-harm and suicide-related content, while also addressing broader risks to young people's mental health and wellbeing.
- Such steps might include further limiting the volume of self-harm- or suicide-related content allowed on platforms, and adapting the algorithms to reduce the frequency with which it is delivered to users, even if they have previously searched for it. Steps to reduce the potential harms associated with exposure may also include ensuring that all self-harm- or suicide-related content is accompanied by educational and/or helpline information
- Develop evaluation approaches for online safety measures (with mechanisms for meaningful youth participation) to ensure regulatory approaches are evidence-based and reflect young people's experiences and needs. This is consistent with recommendation 6 of the *Joint Select Committee on Social Media and Australian Society* final report.(22)
- The legislation should also make clear how compliance will be monitored and enforced, and these mechanisms should be strong. We welcome clarity about the range of enforcement options available to the Australian Government to enforce the legislation and how these will be made transparent to stakeholders and the broader community.

This submission was written on the lands of the Wurundjeri people of the Kulin Nation. Orygen acknowledges the Traditional Owners of the lands we are on and pays respect to their Elders past and present. Orygen recognises and respects their cultural heritage, beliefs and relationships to Country, which continue to be important to the First Nations people living today.

REFERENCES

1. eSafety Commissioner. Recommender systems: position paper. Canberra: Australian Government; 2026.
2. Chhabra J, Pilkington V, Benakovic R, Wilson MJ, La Sala L, Seidler Z. Social Media and Youth Mental Health: Scoping Review of Platform and Policy Recommendations. J Med Internet Res. 2025;27:e72061.
3. Robinson J, Thorn P, McKay S, Richards H, Battersby-Coulter R, Lamblin M, et al. The steps that young people and suicide prevention professionals think the social media industry and policymakers should take to improve online safety. A nested cross-sectional study within a Delphi consensus approach. Frontiers in Child and Adolescent Psychiatry. 2023;Volume 2 - 2023.
4. Hamilton JL, Torous J, Szlyk HS, Biernesser C, Kruzan KP, Jensen M, et al. Leveraging Digital Media to Promote Youth Mental Health: Flipping the Script on Social Media-Related Risk. Curr Treat Options Psychiatry. 2024;11(2):67-75.
5. O'Sullivan S, van Berkel N, Kostakos V, Schmaal L, D'Alfonso S, Valentine L, et al. Understanding What Drives Long-term Engagement in Digital Mental Health Interventions: Secondary Causal Analysis of the Relationship Between Social Networking and Therapy Engagement. JMIR Ment Health. 2023;10:e44812.
6. Cooper A, Brisbane, M. & Mithieux, T. . A digital duty of care for Australia. 2026.
7. Orygen. Social by design: harnessing the benefits of social media for suicide prevention: briefing pack. Parkville, VIC.
8. Cooper C, Louise La Sala, Pamela Patrick, lefteris patlamazoglou, Bridget Kenny, Caroline X Gao, Michelle Lamblin, et al. What are the self-harm and suicide-related social media experiences of LGBTQ+ young people? A national survey comparing LGBTQ+ and non-LGBTQ+ young Australians. 2026.
9. O'Connor. Hatescape: An In-Depth Analysis of Extremism and Hate Speech on TikTok. Institute for Strategic Dialogue; 2021.
10. Sukunesan S. 'Anorexia coach': sexual predators online are targeting teens wanting to lose weight. Platforms are looking the other way 2021 [Available from: theconversation.com/anorexia-coach-sexual-predators-online-are-targetingteens-wanting-to-lose-weight-platforms-are-looking-the-other-way-162938.
11. Giuseppe Logrieco MM, Marco Roversi, & Alberto Villani. The Paradox of Tik Tok Anti-Pro-Anorexia Videos: How Social Media Can Promote Non-Suicidal Self-Injury and Anorexia. International Journal Of Environmental Research And Public Health. 2021;18(3), 1041.
12. Safe Steps. A National Response for Young People: Support for young people worried about family, domestic and intimate partner violence. 2025.
13. A Duncan. AMCT. Family and domestic violence in Australia 2025: is enough progress being made? Federal Election Research Brief #2/25. Bankwest Curtin Economics Centre 2025.
14. Suicide Prevention Australia. AI and digital safety. Sydney, NSW: Suicide Prevention Australia; 2025.
15. Australian Institute of Health and Welfare. Family, domestic and sexual violence: Children and young people 2026 [Available from: https://www.aihw.gov.au/family-domestic-and-sexual-violence/population-groups/children-and-young-people#changed.
16. Dragiewicz M OLP, Ackerman J, Bond C, Foo E, Young A, Reid C. . Children and technology-facilitated abuse in domestic and family violence situations: Full report. Canberra, ACT: eSafety Commissioner, Australian Government; 2020.
17. Dennehy R, Meaney S, Cronin M, Arensman E. The psychosocial impacts of cybervictimisation and barriers to seeking social support: Young people's perspectives. Children and Youth Services Review. 2020;111:104872.
18. Holly L, Vishwanath P, Redlich C, Lazëri L, Azzopardi-Muscat N. Informing Policy Responses to Digital Technology Use and Youth Mental Health and Well-Being: Scoping Review. Online J Public Health Inform. 2026;18:e90160.
19. Nagata JM, Memon Z, Huang O, Moreno MA. Adolescent Health and Generative AI-Risks and Benefits. JAMA Pediatr. 2026;180(1):7-8.
20. eSafety Commissioner. Talking to machines: Children's experiences with AI assistants and companions. Australian Government; 2026.
21. Fraser H CR. There are risks when children use AI chatbots. Does Australia's new 'digital duty of care' address them? 2026 [Available from: https://doi.org/10.64628/AA.gxq4pde4c.

22. Joint Select Committee on Social Media and Australian Society. Social media: the good, the bad, and the ugly – final report. Canberra: Commonwealth of Australia 2024.