



IMPACTS OF ONLINE REGULATION

Summary

- The legislative and regulatory framework contained in the *Online Safety Act 2021* (Cth) (the Online Safety Act) generates risks for young people and marginalised communities by failing to protect sexual and reproductive health education, drug harm reduction information and sex worker support resources from filtering or removal.
- Filters and algorithmic ranking systems already overcapture sexual and reproductive health education on STIs, HIV, contraception, consent, LGBTIQ+ issues; and drug-related harm reduction information.¹⁻⁶ Australia's online regulatory system will result in further overcapture, blocking young people and adults from critical health information.⁷
- The Online Safety Act's regulatory framework imposes high penalties for tech companies that fail to remove or age-gate content, but fails to safeguard against over-removal. There are no statutory or regulatory complaints mechanisms, and no obligations for platforms to create accessible, transparent and responsive appeals processes. This imbalance encourages censorship of legitimate and lawful content.
- Mandated age-assurance mechanisms demonstrate inconsistent accuracy, and platforms are incentivised to collect excessive personal data due to enforcement pressures and liability risks.
- Sex worker advertising platforms and community resources will likely be put behind age-gates, driving away clients, reducing income and undermining sex worker safety strategies.

Recommendations

- Replace Australia's current regulatory framework with one that:
 - centres access to accurate information for all ages
 - protects user privacy
 - prioritises digital participation and protects marginalised communities from digital discrimination
 - promotes consent in platform design, allowing users control over what content they see and choice in using engagement metrics, push notifications, recommender systems and other platform features impacting user experience.
- Establish a mechanism within the Office of the eSafety Commissioner that addresses the blocking or filtering of content that should not be captured, enabling organisations to submit cases for escalation directly with tech platforms.
- Establish an oversight mechanism that includes non-government organisations, the eSafety Commissioner and the Department of Communications; ensuring meaningful participation from sex workers and other marginalised communities including representatives from LGBTIQ+, youth, First Nations and disability rights organisations to surface unintended barriers.
- Ensure that audits required by the eSafety framework, such as platform compliance reports, are made public, with avenues to correct any identified overreach.
- Amend eSafety regulations to explicitly protect sexual and reproductive health education, drug-related harm reduction, and sex worker advertising and support resources from deranking, filtering, and age-gating.

Background

The Online Safety Act and associated regulatory framework empowers the eSafety Commissioner to register 'co-regulatory industry codes of practice' or implement standards for all online services, and grants the Commissioner broad-ranging powers to proactively investigate, issue warnings, delist or remove online content.

The codes and standards impose broad requirements on technology products and services to implement preemptive measures to detect, remove, or restrict access to content for all Australian end-users. These requirements apply to:

- search engines (e.g. Google)
- hosting services (e.g. GoDaddy)
- social media platforms (e.g. Instagram)
- messaging apps (e.g. WhatsApp)
- Internet Service Providers (e.g. Optus)
- app distribution services (e.g. Apple App Store)
- device manufacturers (e.g. Apple).

The phase 2 codes mandate age assurance systems to restrict access to 'class 1C' and 'class 2' materials, which includes content with 'high-impact' depictions of nudity, real and simulated sexual activity, violence, drug use, language, and 'themes' - a term encompassing suicide, racism, family breakdown, death, serious illness, and drug dependency.

 **Table 1: definitions of class 1C and class 2 material**

CLASS	ONLINE MATERIAL
1C	Material that describes or depicts specific fetish practices or fantasies.
2A	Other sexually explicit material that depicts actual (not simulated) sex between consenting adults.
2B	▪ Material which includes realistically simulated sexual activity between adults. ▪ Material which includes high-impact nudity. ▪ Material which includes high-impact sex, nudity, violence, drug use, language and themes. 'Themes' includes social Issues such as crime, suicide, drug and alcohol dependency, death, serious illness, family breakdown and racism.

Concerningly, the term 'theme's in class 2B is expansive. It potentially encompasses nearly any commentary or discussion on social issues: precisely the content young people seek online when school-based education is inadequate. The NSW Parliament's *Inquiry into Harmful Pornography* found that 'current sexuality education is not meeting the needs of a significant number of young people, who turn to pornography to fill gaps in their sexuality education. This gap is particularly pronounced for LGBTIQ+ young people and young people with disability'.⁸

The overcapture problem

Sexual and reproductive health promotion often includes frank discussions and images, such as diagrams of anatomy, condom demonstrations, depictions of sexual acts. Empirical studies show that automated and human moderators routinely misclassify such content as pornography. For example, algorithms have flagged innocuous health illustrations or cartoons as 'adult' content.⁹



Young people routinely ‘turn to the internet for sexual health resources,’ but are often met with ‘misinformation and roadblocks’ from big tech’s content filters.¹ Large platforms (e.g. YouTube, Meta) censor and deplatform sexual and reproductive health education and LGBTIQ+ sexual health discussions by incorrectly moderating them as prohibited adult content.¹⁻⁶ Content moderation systems also perpetuate bias that leads to women’s and queer people’s bodies and health promotion being censored more heavily than men’s.¹⁰⁻¹²

The Center for Intimacy Justice report, *The Digital Gag: Suppression of Sexual and Reproductive Health on Meta, TikTok, Amazon, and Google* documented systematic suppression of sexual and reproductive health information across major platforms, finding that nonprofit organisations, creators, and small businesses routinely had sexual health ads and content removed or had accounts suspended without clear explanation.⁶ In 2025, it was reported that Google had blocked advertisements for an Australian government-funded STI campaign as ‘adult content’ on 12 occasions, and the organisation running the campaign had to contact Google on 10 different occasions.¹³

At its presentation to the 27th Congress of the World Association for Sexual Health, UNESCO analysed content moderation policies across seven major platforms and identified three critical absences:

- **Absence of positive and enabling policies:** frameworks focus on punitive removal rather than clear guidance protecting health content.
- **Absence of distinction:** ambiguity leads to broad enforcement where evidence-based and educational content is treated as an anomaly.
- **Absence of pleasure:** limited or ambiguous guidelines around healthy and positive aspects of sexuality and reproductive health.

UNESCO explicitly warned that automated and human moderation systems ‘pathologise and automatically sexualise legitimate educational content,’ particularly when it relates to gender, sexuality, or pleasure.⁴

Impact on sex worker advertising and safety

Advertising for in-person sex work services are also at risk of being censored and age-gated due to the vagueness of terms within the regulations, or misinterpretation by service providers. Mandating age assurance technologies for access to on advertising sites will dramatically reduce client traffic. International evidence warns of the consequences: after the US passed FOSTA/SESTA in 2018 and forced the closure of Backpage.com, most sex workers’ incomes fell and their working conditions worsened.¹⁴ It also significantly reduced sex workers’ ability to implement safety strategies, including pre-screening of clients and the sharing of health and safety information.^{15,16}

Whilst the purpose and intent of FOSTA-SESTA differs from the Online Safety Act’s regulatory framework, both mandate increased compliance measures, creating a “chilling effect” on online services’ willingness to host sex worker voices and content.¹⁵⁻¹⁷ Platforms’ fear of noncompliance, or overenthusiastic enforcement, could see any discussion of sex work classed as ‘high-impact themes,’ regardless of whether the content is sexually explicit or actually violates any terms of service.

Digital access and equity

Internet access is indispensable in modern life. The Online Safety Act statutory review itself noted that ‘choosing not to be online is no longer a practical option for most Australians’ and that ‘digital exclusion is increasingly a driver of inequality.’¹⁸

Marginalised groups – including rural youth, Indigenous communities, people with disability, or those in abusive homes – often **rely on online channels for health support and information**. Digital health researchers find that technology-based interventions ‘can help to address inequities in sexual health outcomes’ by reaching diverse and underserved adolescents.¹⁹ This is equally true for sex workers, who use internet technologies for advertising, safety and community connection. Censoring or restricting broad categories of content thus risks undermining health promotion, peer education and workplace organising, entrenching disadvantage within our communities.



UK Online Safety Act: a predictor of Australia's future

Australia's online regulatory framework closely mirrors the United Kingdom's Online Safety Act (OSA), which became operational in July 2025. The UK experience demonstrated how these laws function in practice, not as targeted child protection measures, but as systemic censorship engines that suppress legal content while creating new privacy and security vulnerabilities.

Reporting has documented numerous examples of content overcapture impacting UK-based internet users since the OSA came into effect, with automated filters blocking content about Palestine, Ukraine, and even parliamentary debates on serious issues like grooming gangs, catching legitimate journalism and political discourse in filters intended to 'protect children.'²⁰⁻²⁴

While the OSA does not expressly mandate censorship of content related to protests, conflict or sexuality, platforms have erred on the side of removal, creating chilling effects disproportionately impacting UK-based sex workers and LGBTIQ+ people.²⁵ Most concerning for health promotion, online forums related to gender transition, sexual and reproductive health, sexual assault, self-harm support and smoking cessation been subject to age-gating as a consequence of the OSA.^{20,22}

The UK experience also demonstrates the privacy vulnerabilities embedded in poorly-designed platform regulation. Index on Censorship has warned against the OSA's power to mandate platforms' use of 'accredited technology' to scan private messages, ostensibly for child sexual exploitation and terrorism material, but with potential for expansion.²⁵ This undermining of security design mirrors the enforcement pressures building in Australia's framework, creating surveillance infrastructure that, once established, can be repurposed beyond its original intent.²⁶

Australia's online regulatory framework replicates the UK OSA's structural flaws without learning from its documented failures. Unless the government establishes explicit carve-outs for health education, creates accountability mechanisms for overcapture, and ensures meaningful participation from affected communities in enforcement decisions, Australia will replicate the UK's mistakes.

Regulatory abdication

When platforms face massive penalties for under-removal but no consequences for over-removal, they will always err on the side of silence. The voices lost are sex workers, sex educators, LGBTIQ+ people and advocates, harm reduction workers, and peer-led community organisations whose expertise is being erased from the digital public square.

Regulatory guidance produced by the eSafety Commissioner's acknowledges that 'it is unlikely that sexual health education content, information about sexuality and gender, or health and safety information about drug use and sex' would be deemed illegal or restricted.²⁷ However, **'unlikely' is not guaranteed**: with no statutory exemption, any ambiguous image or term could be incorrectly flagged for an age gate or takedown. The stated position of the Department of Communications is that 'it is ultimately a matter for industry about how they implement the codes and what steps they take to address 'over capture' of lawful material.' (A Hyles, Acting First Assistant Secretary, Digital Platforms, Safety and Classification Division, correspondence ref: MC25-016477, 16 September 2025).

While the Online Safety Act's regulatory framework mandates that platforms must have clear and accessible complaints mechanisms to report 'harmful' content, it does not:

- contain any mechanisms for creators or health organisations to appeal wrongful content removal or shadowbanning
- create any obligations for platforms to facilitate meaningful appeals processes
- impose any penalties on platforms for the mis-identification of content and subjecting it to control mechanisms.

This approach transfers responsibility for protecting health information, political speech and digital participation for marginalised communities entirely to platforms, who face massive financial penalties (up to \$49.5 million per breach) for failing to filter or remove content, and no penalties for over-filtering.



This creates a powerful incentive asymmetry favoring removal or restriction, and leaves health promotion organisations and marginalised communities beholden to opaque, inaccessible and ineffective platform appeals mechanisms. Research on platform appeal mechanisms reveals systemic dysfunction: appeals processes are slow, resource-intensive, and often provide no meaningful recourse for wrongfully removed content, particularly for resource-constrained health promotion organisations, LGBTIQ+ advocacy groups, and other nonprofits.²⁸

Age assurance's failures

Age assurance mechanisms mandated by the regulatory framework demonstrate fundamental technical and privacy failures. The Australian Government's Age Assurance Technology Trial (AATT) found that no solutions were 'guaranteed to be effective':

- document-based verification excludes people without identity documents
- biometric age estimation is probabilistic and unsatisfactorily inaccurate for users close to age thresholds (e.g. 19 year-old users verifying that they are over the age of 18)
- age inference is based on behavioural patterns and risks algorithmic bias.²⁹

Excessive data collection: The AATT also found that, despite regulatory guidance, service providers were building systems enabling regulators and law enforcement to 'retrace the actions taken by individuals to verify their age,'²⁹ unnecessarily expanding data collection and retention beyond what was required.

Despite legislation preventing government ID being the *only* mechanism of age assurance, the inaccuracy of other age assurance mechanisms means that some individuals will be forced to use government ID to verify their age. All age assurance systems require the collection and storage of highly sensitive personal information, including government IDs, biometric data, and payment information. Recent breaches demonstrate the vulnerability:

- Tea app breach (July 2025): Exposed over 70,000 images (including 13,000 verification selfies and government-issued IDs), as well as the locations of 33,00 users via the online message board 4chan.³⁰
- Discord data breach (October 2025): Exposed over 68,000 Australian Discord users age-verification information, including government-ID photos, after a third-party service was compromised.³¹

While the eSafety Commissioner's Social Media Minimum Age Regulatory Guidance states that retaining 'information about individual end-users, [and] their unique age assurance checks or outcomes is not necessary' to demonstrate compliance,³² the industry codes and standards do not contain similar guidance. As acknowledged by the AATT, this creates a risk of platforms and/or age assurance businesses retaining individual data for actual or perceived compliance reporting requirements.

References

- 1 Borrás Pérez P. Facebook doesn't like sexual health or sexual pleasure: big tech's ambiguous content moderation policies and their impact on the sexual and reproductive health of the youth. *Int J Sex Health*. 2021;33(4):550-554. <https://doi.org/10.1080/19317611.2021.2005732>
- 2 Williams J. Deplatforming sex education on Meta: sex, power, and content moderation. *Media Int Aust*. 2023;194(1):36-52. <https://doi.org/10.1177/1329878X231210612>
- 3 Gomes AB, Sultan A. Problematizing content moderation by social media platforms and its impact on digital harm reduction. *Harm Reduct J*. 2024;21(1):194. <https://doi.org/10.1186/s12954-024-01104-9>
- 4 Chau K, UNESCO. Algorithms and access: how content moderation shapes digital sexuality education. Presented at: 27th Congress of the World Association for Sexual Health; 16-19 Jun 2025; Brisbane, Australia. Accessed 30 Sep 2025. <https://airdrive.eventsair.com/eventsairaueprod/production-ashm-public/240b5438f4414d9e9a54a7834112c133>
- 5 Garwood-Cross L, Light B, Cooper-Ryan AM, Vasilica C. Sex education against the algorithm: the algorithmically enforced deplatforming of YouTube sex edutainment. *J Gend Stud*. 2025;34(8):1-13. <https://doi.org/10.1080/09589236.2024.2374752>



- 6** Rotman J, Wersbe C, Saud M, Owens J. *The Digital Gag: Suppression of Sexual and Reproductive Health on Meta, TikTok, Amazon, and Google*. Center for Intimacy Justice; 2025. Accessed 5 Dec 2025. <https://docsend.com/view/emzyirq6hfatmx2a>
- 7** Woodley G, Albury K, Stardust Z. Sexual health info online is crucial for teens. Australia's new tech codes may threaten their access. *The Conversation*. 29 May 2025. <https://doi.org/10.64628/AA.jnkqj7hm4>
- 8** Legislative Council Standing Committee on Social Issues. *Impacts of Harmful Pornography on Mental, Emotional, and Physical Health*. Parliament of New South Wales; Oct 2025. Accessed 5 Dec 2025. <https://www.parliament.nsw.gov.au/committees/inquiries/Pages/inquiry-details.aspx?pk=3067#tab-reportsandgovernmentresponses>
- 9** Matsakis L. Tumblr's porn-detecting AI has one job—and it's bad at it. *Wired*. 3 Dec 2018. Accessed 16 May 2025. <https://www.wired.com/story/tumblr-porn-ai-adult-content/>
- 10** Joseph C. Instagram's murky shadow bans just serve to censor marginalised communities. *The Guardian*. 9 Nov 2019. Accessed 5 Dec 2025. <https://www.theguardian.com/commentisfree/2019/nov/08/instagram-shadow-bans-marginalised-communities-queer-plus-sized-bodies-sexually-suggestive>
- 11** Mauro G, Schellmann H. There is no standard: investigation finds AI algorithms objectify women's bodies. *The Guardian*. 8 Feb 2023. Accessed 5 Dec 2025. <https://www.theguardian.com/technology/2023/feb/08/biased-ai-algorithms-racy-women-bodies>
- 12** CensHERship. *Censorship revealed: the impact of digital suppression and censorship of women's health*. In partnership with The Case for Her. Jun 2025. Accessed 5 Dec 2025. <https://www.censhership.co.uk/white-paper>
- 13** May N. Google blocked STI awareness ads as adult content, Senate estimates hears. *The Guardian*. 9 Oct 2025. Accessed 5 Dec 2025. <https://www.theguardian.com/australia-news/2025/oct/09/google-blocked-sti-awareness-ads-as-adult-content-senate-estimates-hears>
- 14** Blunt D, Wolf A. Erased: the impact of FOSTA-SESTA and the removal of Backpage on sex workers. *Anti Traffick Rev*. 2020;14:117-121. <https://doi.org/10.14197/atr.201220148>
- 15** Moraff L. How online censorship harms sex workers and LGBTQ communities. *ACLU*. 27 Jun 2022. Accessed 16 May 2025. <https://www.aclu.org/news/civil-liberties/how-online-censorship-harms-sex-workers-and-lgbtq-communities>
- 16** Albert K, Armbruster E, Brundige E, Denning E, Kim K, Lee L, Ruff L, Simon K, Yang Y. *FOSTA in Legal Context*. SSRN. 30 Jul 2020. <https://doi.org/10.2139/ssrn.3663898>
- 17** Vark C. Sex, money, and politics: big tech made FOSTA-SESTA global for sex workers. *Am Univ Bus Law Rev*. 19 Apr 2023. Accessed 16 May 2025. <https://aublr.org/2021/10/sex-money-politics-big-tech-made-fosta-sesta-global-for-sex-workers/>
- 18** Department of Infrastructure, Transport, Regional Development, Communications and the Arts. *Statutory Review of the Online Safety Act 2021*. 29 Apr 2024. Accessed 16 May 2025. <https://www.infrastructure.gov.au/department/media/publications/statutory-review-online-safety-act-2021-issues-paper>
- 19** Borji-Navan S, Maleki N, Keramat A. Efficacy of digital health interventions used for adolescent's sexual health: an umbrella review. *Health Sci Rep*. 2024;7(12):e70212. <https://doi.org/10.1002/hsr2.70212>
- 20** Grieg J. The UK's new online safety laws are already a disaster. *Dazed*. 30 Jul 2025. Accessed 5 Dec 2025. <https://www.dazeddigital.com/life-culture/article/68340/1/britain-uk-new-online-safety-laws-are-already-disaster-internet-censorship>
- 21** Davies N, Sardarizadeh S, Murphy M. Tech giants blocking some Ukraine and Gaza posts under new online rules. *BBC*. 31 Jul 2025. Accessed 5 Dec 2025. <https://www.bbc.com/news/articles/cj3l0e4vr0ko>
- 22** Gallagher M. Age-gated: how the Online Safety Act is distorting digital activism. *Byline Times*. 8 Aug 2025. Accessed 5 Dec 2025. <https://bylinetimes.com/2025/08/08/age-gated-how-the-online-safety-act-is-distorting-digital-activism/>



- 23** Lorenz T. The UK's Online Safety Act is a licence for censorship – and the rest of the world is following suit. *The Guardian*. 9 Aug 2025. Accessed 5 Dec 2025. <https://www.theguardian.com/commentisfree/2025/aug/09/uk-online-safety-act-internet-censorship-world-following-suit>
- 24** Scarlett M. Online Safety Act: protect kids is pretext for rights erosions. *Open Democracy*. 8 Oct 2025. Accessed 5 Dec 2025. <https://www.opendemocracy.net/en/beyond-trafficking-and-slavery/online-safety-act-protect-the-children-is-pretext-for-rights-erosions/>
- 25** Steinfeld J. The Online Safety Act risks making everyone less safe. *Index on Censorship*. 15 Aug 2025. Accessed 5 Dec 2025. <https://www.indexoncensorship.org/2025/08/the-online-safety-act-risks-making-everyone-less-safe-online/>
- 26** Garcia RT. Britain calls it safety. It is censorship. *Al Jazeera*. 6 Nov 2025. Accessed 5 Dec 2025. <https://www.aljazeera.com/opinions/2025/11/6/britain-calls-it-safety-it-is-censorship>
- 27** eSafety Commissioner. *Online Content Scheme Regulatory Guidance ESC RG*. 4 Jan 2025. Accessed 5 Dec 2025. https://www.esafety.gov.au/sites/default/files/2025-01/Online-Content-Scheme-January2025_0.pdf
- 28** Are C. Dysfunctional appeals and failures of algorithmic justice in Instagram and TikTok content moderation. *iCS*. 2025;28(11):1997-2014. <https://doi.org/10.1080/1369118X.2024.2396621>
- 29** Age Check Certification Scheme. *Age Assurance Technology Trial: Part A - Main Report*. Aug 2025. Accessed 5 Dec 2025. <https://ageassurance.com.au/report/>
- 30** Wakefield J. Tea app: I joined for safety. Then my address was leaked and shared. *BBC*. 23 Aug 2025. Accessed 9 Dec 2025. <https://www.bbc.com/news/articles/ce87rer52k3o>
- 31** Booth R. Hack of age verification firm may have exposed 70,000 Discord users' ID photos. *The Guardian*. 9 Oct 2025. Accessed 9 Dec 2025. <https://www.theguardian.com/media/2025/oct/09/hack-age-verification-firm-discord-users-id-photos>
- 32** eSafety Commissioner. *Social Media Minimum Age Regulatory Guidance*. Sep 2025. Accessed 9 Dec 2025. <https://www.esafety.gov.au/industry/regulatory-guidance#social-media-minimum-age>

