Equality Australia
ABN: 20 609 977 764
262 Liverpool St, Darlinghurst NSW 2010
info@equalityaustralia.org.au
equalityaustralia.org.au
Ph: (02) 7208 7922

30 October 2026

By email

digitaldutyofcare@communications.gov.au

STRENGTHENING THE DIGITAL DUTY OF CARE BILL

Thank you for the opportunity to make a submission on the Online Safety Amendment (Digital Duty of Care) Bill 2026.

Our principal concern is that there remains a significant gap in the Bill from an LGBTIQ+ safety and harm-prevention perspective. This is due to the unreasonably high threshold for a safe online environment for all persons set by clause 25C, and particularly the narrow and exhaustive approach to 'seriously harmful material and conduct' in 25D.

ONLINE HARMS TARGETING LGBTIQ+ PEOPLE

Online hate against LGBTIQ+ people is a significant and ongoing experience for our communities. According to eSafety, the LGBTIQ+ community experiences online hate speech at more than twice the rate of the general population in Australia.[1] Research conducted by ANROWS in 2022 found that 3 in 4 (72.7%) lesbian, gay and bisexual people had experienced tech-based abuse.[2]

Research commissioned by eSafety in 2021 found that 8 in 10 LGBTIQ+ teens had seen potentially harmful content in the past year, including sites showing hate messages (60% as compared with 40% in the broader population). Similarly to the study on adults, LGBTIQ+ teens had experienced hate speech at twice the national average, and even more concerningly 1 in 4 had been threatened with physical harm.[3]

We have observed a concerning proliferation of social media posts, blogs, and online forums that incite hatred and violence against LGBTIQ+ people. These include posts that:

- call for the genocide or for the 'end' of trans people
- imply or encourage that trans people should be assaulted, hanged or even subject to genocide (i.e. incorporating imagery of a wood chopper, or a hanging man)
- liken the LGBTIQ+ movement to the Ku Klux Klan or Nazis, or fascism generally including concepts from 1984 such as 'Newspeak' or 'Wrongthink'

---

[1] eSafety Commissioner (Cth), *Fighting the Tide: Encounters with Online Hate among Targeted Groups* (Report, February 2025) 19; see also eSafety Commissioner, 'Online Hate and Discrimination' (Web Page, 27 October 2026) <https://www.esafety.gov.au/lgbtiq/learning-lounge/dealing-with-online-abuse/online-hate-discrimination>.

[2] ANROWS, Technology-facilitated abuse: National survey of Australian adults' experiences (Research Report, December 2022) 23.

[3] eSafety Commissioner, Tipping the balance: LGBTIQ+ teens' experiences negotiating connection, self-expression and harm online (Aussie Kids Online Report, June 2024) 12.

- portray gay men as paedophiles or sexual predators
- refer to trans people as 'sex deceptionists' or claim they should be institutionalised
- describe trans women as predators, paedophiles, or inherently evil.

These forms of vilification target both the broader LGBTIQ+ community and specific individuals, including community members as well as high-profile figures such as community advocates, politicians and athletes. Once a person becomes the focus of online hate, it can spiral into a sustained campaign, sometimes involving hundreds or thousands of posts that fixate on their identity, appearance, and personal life. We are aware that this has caused major anxiety and fear in some individuals, who become scared to even leave their own homes.

CONTEXT FOR THE DIGITAL DUTY OF CARE

It is important not to lose sight of the context of these reforms, and how distinct it is from other typical contexts in which harmful speech is regulated. The relevant question in relation to this Bill is whether the online environment creates or amplifies a foreseeable risk of harm and whether reasonable steps could be taken to prevent or minimise that risk.

Criminal law appropriately sets a very high threshold, given that criminal offences can result in significant consequences including loss of liberty and a criminal record. Civil law also imposes significant consequences, including liability for compensation. The threshold for complaints under the Online Safety Act also must remain reasonably high, since it can lead to removal of content produced by individual users, and therefore raises some legitimate concerns regarding free expression of individuals.

The digital duty of care serves a very different purpose. It is not intended to determine whether a single piece of content is unlawful or impose direct liability on an individual for their conduct online. Rather, it is a preventative regulatory framework directed at the very well-resourced services through which online harms occur.

A service can reasonably be expected to identify and address a foreseeable risk of harm even where no individual user has committed a criminal offence. However, the exposure draft sets a very high threshold for material to be considered 'seriously harmful' to adults in the context of hate speech, by limiting the relevant conduct to existing Commonwealth criminal offences (see clauses 25D(k)-(o)).

HATE AGAINST TARGETED GROUPS

As we have previously expressed, we are concerned that clause 25D does not expressly capture targeted attacks against LGBTIQ+ and other communities and therefore fails to respond adequately to an escalating and deeply concerning situation for our communities.

Indeed, this could be *counterproductive* – sending a message that the expectations for responding to material and conduct falling within the scope of the duty is even lower than current legal obligations under civil vilification law that platforms must already comply with.

Social media companies are unable or unwilling to moderate unlawful and/or harmful content and, in recent times, have actively and publicly stepped back from proactive measures.

For example, since X Corp (formerly Twitter) was purchased by Elon Musk in October 2022, the frequency of posts containing homophobic, transphobic and racist slurs has significantly increased. An analysis found that there has been a doubling of 'likes' on posts containing hate speech and increased

engagement on those posts.[4]

Other platforms including Meta have also publicly stated that they have substantially eased their moderation, specifically by removing restrictions on topics like gender identity, no longer proactively detecting less serious policy violations, and relying more heavily on user reports before taking action.[5] These changes were examined by the Royal Commission on Antisemitism and Social Cohesion in July 2026, which heard that Meta had moved towards a largely reactive model of enforcement, relying on users to report hateful content rather than proactively detecting and removing it. For a 3-month period in 2025, Meta removed 5.18 million posts, but by July – August 2025, they only removed 1.2 million items of hate speech.[6]

Therefore, unless the threshold is sufficiently high, it may in fact reinforce this stance.

On the other hand, we recognise that a threshold such as section 18C of the *Racial Discrimination Act 1975* may not be suitable in this context either. In our view, an appropriate middle ground could be reached to address the kind of seriously harmful speech that falls just short of criminal conduct, where it is directed towards either individuals or communities.

Recognising targeted hate as an online safety risk

The Bill should expressly recognise online attacks based on protected characteristics as a category of seriously harmful material and conduct. This would also give effect to the recommendations of the statutory review of the *Online Safety Act*, particularly Recommendation 5, which identified attacks based on protected characteristics as a harm requiring attention, and Recommendations 21 and 22 concerning online hate and cumulative harm.

We note that certain types of harmful speech might be captured under clause 25D(1)(m). Section 474.17 of the Criminal Code already makes using a carriage service to menace, harass or cause offence a Commonwealth offence, and there are also provisions under Div 80 of the Code that make it a crime to advocate or threaten physical force or violence against targeted groups.

However, the gap created by the current drafting is the substantial amount of hate speech online that may fall short of the threshold for criminal liability, but may nevertheless lead to serious harm, particularly where it is repeated, coordinated, amplified or directed at a community.

We have observed some inconsistency in clause 25D. The provision expressly recognises as seriously harmful material that encourages, promotes, incites or instructs individuals to cause harm to, or engage in cruelty against, animals. It also recognises material that encourages or promotes suicide, self-harm and sadistic online exploitation. These are not direct physical harms themselves but are presumably included because they could encourage or normalise harmful activities.

It is difficult to see in this context why these harms would be serious enough to meet the threshold for 'seriously harmful' material, while vilifying and dehumanising speech directed at marginalised communities is not included, despite evidence that this form of hate can cause serious psychological harm to individuals and communities, contribute to the conditions in which offline harm occurs, and

---

[4] Daniel Hickey et al, 'X under Musk's leadership: Substantial hate and no reduction in inauthentic activity' (2025) 20(2) *PLoS ONE* 1, 1.
[5] Meta, 'More Speech and Fewer Mistakes' (News Release, 7 January 2025) <https://about.fb.com/news/2025/01/meta-more-speech-fewer-mistakes/>. Meta, 'More Speech and Fewer Mistakes' (News Release, 7 January 2025) <https://about.fb.com/news/2025/01/meta-more-speech-fewer-mistakes/>.
[6] Australian Broadcasting Corporation, 'Meta Rejects Claims Antisemitism Increased on Its Platforms Following Relaxation of Content Moderation Rules' (ABC News, 6 July 2026) <https://www.abc.net.au/news/2026-07-06/royal-commission-meta-rejects-antisemitism-increase-claims/106882340>.

normalise prejudice in ways that can escalate into discrimination, harassment and violence.

Proposed amendment to clause 25D

We suggest that 25D could be amended to include an additional provision:

> *material or conduct that vilifies*[7] *or dehumanises*[8] *a targeted group, distinguished by race, religion, sex, sexual orientation, gender identity, intersex status, disability, nationality, national or ethnic origin, where it is reasonably likely that the material or conduct may to lead to serious harm.*

Or alternatively, a narrower approach could be taken, only including the well-established concept of 'hatred':

> *material or conduct that encourages, promotes, incites or instructs individuals to hatred*[9] *of a targeted group, distinguished by race, religion, sex, sexual orientation, gender identity, intersex status, disability, nationality, national or ethnic origin, where it is reasonably likely that the material or conduct may to lead to serious harm.*

Linking the provision to Division 80 would also ensure that the DDOC recognises the characteristics already identified by Commonwealth criminal law for consistency. We have intentionally not included 'political opinion' in this proposed provision. Regulating material on the basis of political opinion raises distinct concerns in the context of online speech and may be incompatible with the implied freedom of political communication.

In our suggestion above, 'serious harm' would also be retained as a key element, which is already defined to include serious physical or psychological harm and distress, but *not* 'mere ordinary emotional reactions such as those of only distress, grief, fear or anger'. This would ensure an appropriately high threshold remains, while addressing targeted harms, consistent with the purpose of the reforms recommended by the statutory review.

In practical terms, the duty holder would not have to identify every piece of vilifying or dehumanising speech online – but instead would need to conduct a risk assessment to identify the extent to which this kind of serious harm is 'reasonably foreseeable' on their platforms, taking into account factors like:

- Nature and seriousness of the material / conduct
- Who are the targets and how many targets are there
- Reach and scale / extent of chance for rapid dissemination

---

[7] By vilify we mean to incite hatred towards, serious contempt for, revolution towards, or severe ridicule of a person or group of persons because of a protected characteristic. This reflects the dominant language used in civil vilification laws across Australian jurisdictions, namely Queensland, NSW, Victoria, ACT.

[8] Dehumanisation involves portraying people, on the basis of a protected characteristic, as lacking human qualities or dignity, including as inherently dangerous, inferior, contaminating or undeserving of empathy or belonging. Dehumanising representations can contribute to an environment in which exclusion, discrimination and violence are more readily accepted against certain groups. Refer to: Australian Muslim Advocacy Network, *Combatting Online Dehumanisation of Minorities* (Policy Brief) <https://www.aman.net.au/policy-brief-combatting-online-dehumanisation-of-minorities/>. While we believe there are benefits to this approach, it hasn't been tested in the law at present.

[9] Hatred in this context means does not mean the emotional feeling of dislike, but refers to prejudice enacted through speech that reinforces the marginalisation of groups in Australia, having regard to the social context in which the speech occurs. See Katharine Gelber, *Speech Matters: Getting Free Speech Right* (University of Queensland Press, 2011) 83; Katharine Gelber, 'Differentiating Hate Speech: A Systemic Discrimination Approach' (2021) 24(4) *Critical Review of International Social and Political Philosophy* 393, 407. There is ample case law interpreting the meaning of the term in Australia and internationally.

- Whether it involves cumulative harm (e.g. pile ons)
- Whether it involves inauthentic activity (e.g. bots).

RECOGNISING COMMUNITY LEVEL HARM

Further, we have identified another concern in the use of the term 'individual' in clause 25D(1)(i), when oftentimes harmful online conduct is directed towards entire communities.

For example, we are aware of a post on X involving an image of a trans flag with a person hanging by a noose with the words 'was/were'.[10] Despite complaints, the post remains online at the time of writing.

In November 2025, Thomas Fordham pled guilty to threatening or inciting violence online saying "We should have a geneside [sic] and kill al [sic] trans people".[11]

The Online Hate Prevention Institute (OHPI) provided a detailed briefing in August 2024, documenting six examples of anti-trans content on social media. In one of these posts, trans women were described as "pedophile trans creatures" in the context of a conspiracy theory that politicians were trying to replace all women.[12]

A separate brief by the OHPI detailed anti-LGBTIQ+ content during Pride Month. One example is an image of a crocodile in the colour of trans flag, with a snake pointing a gun at the crocodile, and the words "Stay the fuck away from our kids, freak."[13]

All of these examples were targeted towards all trans people, not a single individual.

A simple way to address this would be to add "or community":

> (f) ..... by menacing or harassing an individual **or community**.

We do not see any need to define 'community', but alternatively it could be narrowed to characteristics covered by Division 80 of the *Criminal Code* if the above approach is considered to be too broad:

> (f)... by menacing or harassing an individual or a targeted group, distinguished by race, religion, sex, sexual orientation, gender identity, intersex status, disability, nationality, national or ethnic origin.

GENDER-BASED VIOLENCE

We welcome the inclusion of an additional category of harm, not included in the exposure draft Bill, relating to material that encourages, promotes, incites or instructs gender-based violence.

The definition of gender-based violence, which encompasses physical and non-physical violence, harassment, abuse or threats based on gender that result in, or are likely to result in, harm, coercion, control, fear or deprivation of liberty or autonomy, may be broad enough to capture a range of harmful conduct experienced by LGBTIQ+ people online.

This is consistent with eSafety's recognition that "tech-based gendered violence mainly affects women

---

[10] Fraser Anderson (@FraserDAnderson), (Post on X, 2 May 2025) <https://x.com/FraserDAnderson/status/1918215195002061245>.
[11] Adeshola Ore and Jordyn Beazley, 'Man Becomes First to Be Convicted of Hate Speech against Trans People in NSW', *The Guardian* (online, 21 November 2025) <https://www.theguardian.com/australia-news/2025/nov/21/man-becomes-first-to-be-convicted-of-hate-speech-against-trans-people-in-nsw-ntwnfb>.
[12] Online Hate Prevention Institute, 'Vilification and Demonisation of Trans People on Social Media' (Web Page) <https://ohpi.org.au/vilification-and-demonisation-of-trans-people-on-social-media/>.
[13] Online Hate Prevention Institute, 'Anti-LGBTIQ Hate on X (Twitter) during Pride Month' (Web Page) <https://ohpi.org.au/anti-lgbtiq-hate-on-x-twitter-during-pride-month/>.

and girls, and extends to LGBTIQ+ people. This is because it may be based on multiple forms of prejudice, including sexism, misogyny, homophobia, biphobia and transphobia."[14]

However, the legislation and its supporting materials should more clearly recognise that gender-based violence targets LGBTIQ+ people. This would provide greater certainty about the intended scope of the definition and help ensure that harms experienced by our communities are not overlooked in its implementation.

At a minimum, this should be addressed through an explanatory note confirming that gender-based violence includes violence targeting people because of their sexual orientation, gender identity or innate variations in sex characteristics. Alternatively, the definition could expressly recognise violence based on these attributes, while retaining its broader application to other forms of gender-based violence.

We suggest clause 25D(1)(h) is redefined as follows:

> *All forms of physical and non-physical violence, harassment, abuse or threats based on gender, sexual orientation, gender identity, or sex characteristics, including violence based on actual or perceived non-conformity with social or cultural expectations concerning male and female bodies, gender roles, relationships or sexuality, that result in, or are likely to result in, harm, coercion, control, fear or deprivation of liberty or autonomy.*

PRIVATE RIGHT OF ACTION

The digital duty of care as drafted is appropriately focussed on prevention, but we consider more could be done to redress injury after it occurs.

Section 26C applies a civil penalty to a breach of the duty of care. But we would support the addition of a private right of action that ensures that a person who suffers loss or damage because an online service breached its duty can take action for redress, rather than only relying on regulatory action. While compensation should be available as a remedy, the priority in our view is a form of injunctive relief in circumstances a platform fails to address targeted harassment against an individual or an affected community, in breach of the duty of care.

A civil cause of action could provide a means of seeking immediate redress where an online service provider has breached its duty to address a foreseeable risk of harm and that breach has caused personal injury, including psychological injury.

A private right of action could be particularly effective where a person is subjected to a cumulative attack or 'pile-on' across a social media platform. In these circumstances, the harm may arise from the volume, persistence and cumulative effect of multiple posts, comments or interactions, rather than from any single communication considered in isolation. This would recognise that online harm can be cumulative and serious even where individual contributions, considered separately, may not meet the threshold for criminal liability. This approach would be consistent with the recommendations of the Statutory Review of the Online Safety Act 2021, which recognised the need to better address the cumulative harm caused by online hate and to address volumetric attacks as a whole, rather than assessing individual items of abusive material in isolation (Recommendations 22, 23, 24).

Standing should be available to both affected individuals or representative groups where distinguished by race, religion, sex, sexual orientation, gender identity, intersex status, disability, nationality, national or ethnic origin. This could model an approach to representative complaints in Queensland vilification

[14] eSafety Commissioner, 'Gender-Based Violence' (Web Page) <https://www.esafety.gov.au/key-topics/gendered-violence>.

law.[15]

For example, the legislation could provide that:

- a body corporate or unincorporated body whose primary purpose is promoting the interests or welfare of a particular targeted group may bring proceedings;
- the alleged breach must affect, or be likely to affect, people whose interests or welfare are a primary purpose of the organisation;
- the proceedings must be brought in good faith; and
- the court must be satisfied that it is in the interests of justice to permit the proceedings.

APPROVED RESEARCHER DEFINITION

Clause 205B refers to an 'approved researcher' and enables research into online harms in the public interest. We consider this approach may be too narrow if it is limited to Australian universities. Organisations such as the Online Hate Prevention Institute are well-established, properly constituted organisations that are already doing this work effectively, and there may be other community-led organisations that develop this expertise in the future.

We suggest broadening the definition of 'approved researcher' to include appropriately constituted non-profit organisations with demonstrated expertise in monitoring and researching online harms.

We would be happy to discuss any of the above further.

Kind regards

Anna Brown
CEO, Equality Australia

[15] *Anti-Discrimination Act 1991* (Qld) s 134(3)–(5).