



Pathways to address online racism

September 2026

Contents

Summary	2
1. Digital duty of care	3
2. Misinformation and disinformation.....	4
3. Legal protections against online racism	5
4. Fact-checking and counter-speech initiatives	7
Endnotes	8

Summary

Racism makes online environments unsafe and harmful for everyone. Aboriginal and Torres Strait Islander peoples and other communities with lived experience of racism face pervasive racism online, including harassment, hateful speech and the spread of dehumanising and racist ideas.¹ Many community members reported that they experience racism most frequently on social media and in other online environments,² while Aboriginal and Torres Strait Islander adults are twice as likely as non-Indigenous adults to personally experience online hate.³ Online racism often intensifies when issues about racism gain mainstream media and political attention, which amplifies harmful views about communities.⁴

Aboriginal and Torres Strait Islander peoples and other communities with lived experience of racism are disproportionately targeted by online racism and hate speech that can incite hostility, reinforce harmful stereotypes and distort public narratives. The impacts of online racism are compounded when it occurs with intersecting forms of discrimination. These harms are often amplified by misinformation and disinformation which can be used to legitimise racial vilification and even incitement of violence.

Addressing these harms requires a multi-layered, rights-based approach. This includes proactive safeguards to ensure digital platforms assess and mitigate how their systems contribute to harm caused by racism, alongside improved detection of hate speech, measures to limit amplification and accessible redress mechanisms. Fact-checking and counter-speech initiatives (directly challenging racist speech online) also play a critical role in correcting false narratives and promoting a healthy online environment for respectful dialogue that supports anti-racism values, particularly when led by trusted voices from communities with lived experience.

Ultimately, interventions must go beyond reactive moderation. They should form part of a broader strategy that combines platform accountability, public education, inclusive design and community empowerment. This ensures that all people, regardless of race or background, can participate safely and equally in digital life.

Freedom of expression is a cornerstone of democratic society and must be preserved in any response to online racism and hate. A rights-based approach requires striking a careful balance between protecting individuals from racism and safeguarding the ability to challenge ideas, scrutinise policy and engage in robust public discourse. Measures must be proportionate, transparent and accompanied by safeguards. This ensures that efforts to combat harmful content can foster online environments that support critical conversations about anti-racism without inadvertently suppressing diverse viewpoints or stifling open discussion.



1. Digital duty of care

The proposed Digital Duty of Care is a forward-looking regulatory concept that places a legal obligation on digital platforms to take reasonable and proportionate steps to prevent foreseeable harms to users.⁵ It reflects a shift away from reactive, ‘whack-a-mole’ moderation of online content towards a proactive model of platform governance. It is consistent with recommendations from the

Australian Human Rights Commission’s [National Anti-Racism Framework](#) to strengthen reporting mechanisms on digital platforms and introduce a positive duty to prevent racism.

The Australian Government has now committed to introducing an overarching Digital Duty of Care. This will likely require platforms to embed safety by design, conduct regular risk assessments and implement mitigation strategies tailored to their risk profile.⁶ A well-designed Digital Duty of Care could help address persistent harms such as online hate, incitement of violence and algorithmic amplification of hateful content.

This is especially relevant for Aboriginal and Torres Strait Islander peoples and other communities with lived experience of racism, who are often the targets of dehumanising online abuse.⁷ A Digital Duty of Care could require platforms to assess how their systems may contribute to the spread or amplification of racially motivated hate.⁸ It could also require platforms to take proactive steps to reduce these risks, such as:

- Improving detection and removal of foreseeable harms associated with hate speech
- Limiting the virality of harmful content through algorithmic design
- Providing accessible reporting tools and meaningful redress pathways for affected users.⁹

To be effective and rights-respecting, the Digital Duty of Care must be grounded in transparency and accountability, with clear obligations and independent oversight. Where possible, the Digital Duty of Care should be developed and implemented in collaboration with Aboriginal and Torres Strait Islander peoples and other communities with lived experience of racism. At the same time, it must include safeguards to ensure that efforts to reduce harm do not unjustifiably restrict debate and discussion. These protections are essential to maintaining an open and inclusive digital environment where all people can participate freely while being safe from the harms of racism. A safer online environment where racial vilification is reduced or eliminated helps to promote freedom of expression. It enables all people, and especially those who experience racism, to participate more equally in digital life.

By embedding the Digital Duty of Care into law, Australia has the opportunity to lead in creating safer, more inclusive online environments, where all people can participate without fear of abuse, discrimination or exclusion.



2. Misinformation and disinformation

Online racism is one of the most persistent forms of digital abuse targeting Aboriginal and Torres Strait Islander peoples and other communities with lived experience of racism.¹⁰ It often takes the form of dehumanising language and racial slurs that can incite violence or reinforce harmful stereotypes, as well as the promotion of harmful narratives over time. These harms are frequently compounded by misinformation and disinformation which can normalise and enable racism.¹¹

Determining what is ‘misinformation’ and ‘disinformation’ is difficult. The Commission adopts the definitions provided by the Australian Electoral Commission’s Electoral Integrity Assurance Taskforce:

Misinformation is false information that is spread due to ignorance, or by error or mistake, without the intent to deceive.

Disinformation is knowingly false information designed to deliberately mislead and influence public opinion or obscure the truth for malicious or deceptive purposes.¹²

The Commission has consistently emphasised that responses to online racism must be grounded in human rights law and take an anti-racist and systemic approach. This includes understanding how misinformation and disinformation can amplify racist narratives while ensuring strong protections for both the right to equality and non-discrimination and freedom of speech and expression. A multi-faceted approach is essential – one that addresses the structural drivers of harm without empowering any single actor to become the sole arbiter of truth. In the context of misinformation and disinformation, algorithmic and commercial incentives that shape the creation, amplification and monetisation of harmful content can contribute to the spread of racist narratives and other forms of online harm.

A rights-respecting response must include:

- **Clear safeguards for freedom of expression**, ensuring that efforts to reduce harm do not suppress legitimate public debate. Regulation must be clear, proportionate and independently overseen.
- **Targeted action on online racism**, recognising that misinformation and disinformation are often used to legitimise or amplify racism. This includes improving detection of coordinated inauthentic behaviour and algorithmic amplification of hate speech and narratives.
- **Transparency and accountability**, including strengthened reporting obligations for digital platforms on how they moderate content and protect free expression.
- **Collaborative fact-checking** (such as community style notes) allows for a diversity of views and voices to work together and ensure accurate information is promoted without resorting to censorship. Where possible, it is best practice to adequately resource and collaborate with community organisations. For example, community-led racism registers and community-controlled media have expertise, networks and experience countering racist information.



- **Strengthening digital and media literacy**, empowering individuals to critically assess online content, understand algorithmic curation and identify false or harmful narratives.
- **Regulation of high-risk technologies**, such as deepfakes and generative artificial intelligence, to address existing gaps in the legal framework and ensure that the most serious online harms are regulated in precise and targeted ways.

The Commission supports the continuation of measures such as the [Australian Code of Practice on Disinformation and Misinformation](#) as a voluntary, outcomes-based framework that complements broader efforts. However, this is just one measure undertaken by platforms. Further actions in areas such as education and the management of algorithms are also needed as part of a multifaceted approach.

3. Legal protections against online racism

Currently, the legal protection framework against online hate mainly consists of federal laws addressing:

- Racial discrimination and racial vilification (*Racial Discrimination Act 1975* (Cth) (RDA))
- Cyber abuse of adults and bullying of children (*Online Safety Act 2021* (Cth))
- Menacing and extreme forms of speech inciting hatred, violence or supporting terrorism (*Criminal Code Act 1995* (Cth)), and
- Seriously vilifying content (*Broadcasting Services Act 1992* (Cth)).¹³

More recently, the *Combatting Antisemitism, Hate and Extremism (Criminal and Migration Laws) Act 2026* (Cth) introduced higher penalties for hate crime offences and expanded prohibitions against hate symbols.¹⁴

Most states and territories also prohibit hate speech and incitement, as well as some hate symbols, by law.¹⁵

There are, however, several limitations to existing laws. These include the inability to keep pace with technological developments and shifts in different forms of online racism. Hate crime experts have highlighted the unique nature of online racism, where bad actors can adapt methods or wordings of harassment to circumvent moderation or even amplify their message through algorithms, often turning individual hate incidents into a public

harm.¹⁶ Online racism can also occur in indirect ways, such as through the racist curation and promotion of select information or stories to create dehumanising and harmful narratives about targeted communities over time.¹⁷ However, current laws do not cover encounters with these patterns of hateful speech, harassment or narratives that are sometimes not targeted at individuals.¹⁸



Online racism also goes beyond the algorithmic amplification of hate speech. For Aboriginal and Torres Strait Islander communities, online racism can take the forms of the theft and misappropriation of their data (such as images and user data).¹⁹ Ensuring Aboriginal and Torres Strait Islander peoples' ownership and governance of Indigenous data, and embedding Indigenous Data Sovereignty and Governance²⁰ in any response,²¹ is therefore crucial to curbing online racism and enabling Aboriginal and Torres Strait Islander peoples' self-determination.

Hate crime experts and communities have called for urgent action to address limitations in existing responses. This includes:

- Stronger obligations for digital platforms to proactively mitigate harmful content
- Harm-based, community-centric reforms that provide alternatives to imprisonment²² and acknowledge the harm inflicted on both targeted individuals and relevant communities
- Holistic and preventative action to foster a public information environment intolerant of any forms of racism, dehumanisation and discrimination.²³

To foster an anti-racist online environment, the Commission's [National Anti-Racism Framework](#), released in November 2024 following extensive community consultations, recommends various actions. These include stronger legal protections against online hate, embedding Indigenous Data Sovereignty and Indigenous Data Governance in anti-racism action and introducing a positive duty under the RDA. A positive duty would require digital platforms to have policies and procedures in place to prevent racism.

Beyond legislation, a multi-pronged, systemic and comprehensive approach is critical to addressing the roots of online racism. The Framework recommends solutions including:

- Digital platforms developing stronger, transparent protocols to allow users to report and request platforms to remove racist content
- Media organisations adopting guidelines that are grounded in an anti-racist approach to reporting about Aboriginal and Torres Strait Islander peoples and other communities with lived experience of racism, and
- Governments increasing funding for community-led initiatives that take a strengths-based approach to storytelling about Aboriginal and Torres Strait Islander peoples and other communities with lived experience of racism.

Critical to a human rights-based approach to tackling online hate is the principle of participation. Any action to tackle online hate should centre community input and feedback, especially since people from racialised communities are most affected by online racism and are therefore best placed to evaluate the effectiveness of any solutions.



4. Fact-checking and counter-speech initiatives

Fact-checking and counter-speech initiatives are tools in addressing the spread of racist speech and narratives online. Fact-checking helps to expose and correct racist narratives by providing accurate, evidence-based information. When platforms apply fact-check labels or redirect users to credible sources, it may reduce the visibility and impact of racially charged content without resorting to censorship. This is especially important in contexts where hate speech or incitement to violence is falsely presented as opinion. At the same time, fact-checking efforts should be carefully designed to avoid being used to suppress legitimate differences of opinion and ensure that no single actor becomes an unquestioned arbiter of truth.

Counter-speech initiatives that directly challenge racist speech online (particularly when led by trusted voices from affected communities) can play a powerful role in challenging hate and racism, affirming dignity and promoting inclusion. These responses not only correct falsehoods but also model anti-racist dialogue and help build digital resilience, especially among young people who may be encountering hateful speech and narratives for the first time. Where possible, community-led initiatives should be adequately resourced to lead the development and evaluation of counter-speech efforts.

While fact-checking and counter-speech initiatives are often grounded in the principle that 'good speech' will counter 'bad speech' in a free and open exchange of ideas, this ideal does not operate on a level playing field.

In theory, the contest of ideas allows truth to prevail through open debate. However, in practice, communities with lived experience of racism often face disproportionate risks when participating in these online debates. Engaging in counter-speech can expose individuals to further harassment, doxxing or targeted abuse. As a result, the burden of correcting harmful narratives often falls on those most affected by them. A rights-based approach must recognise these asymmetries and ensure that counter-speech efforts are supported, resourced and accompanied by systemic protections against online hate.

Find out more

For more about the National Anti-Racism Framework and to support the vision visit:

www.humanrights.gov.au/anti-racism-framework

or scan the QR code



Acknowledgement

We acknowledge the ongoing strength and leadership of Australia's First Peoples in leading anti-racism efforts since colonisation. The Commission acknowledges the vast emotional and intellectual labour of the community, particularly those with lived experiences of racism, who shared their vision for a National Anti-Racism Framework that takes a transformative approach to addressing racism in Australia.

Illustrations by Bree Buttenshaw for Saltwater People (2024).

Endnotes

- 1 Amanuel Elias, Fethi Mansouri and Yin Paradies, *Racism in Australia Today* (Palgrave Macmillan Singapore, 2021) 228-229.
- 2 Examples of social media and other online environments include social networking sites or applications, online forums, and streaming and gaming platforms. See also, Fiona Allison, Chris Cunneen, Amelia Whyman, Beck Lewis and Ayse Selcuk, *Everywhere I go no matter where, if it's around non-Indigenous people I feel a hate vibe. It feels like I'm being watched* (Annual Report, The Call It Out Racism Register, Jumbunna Institute for Indigenous Education and Research, University of Technology Sydney, 27 November 2025).
- 3 eSafety Commissioner, *Fighting the tide: Encounters with online hate among targeted groups* (Report, February 2025).
- 4 With recent examples including the spike in online racism against Aboriginal and Torres Strait Islander peoples around the Voice Referendum, the rise in anti-Asian racism around the COVID-19 pandemic, and the rise in anti-Palestinian and anti-Arab racism, Islamophobia, and antisemitism that have devastating impacts on communities. See, for example, Australian Human Rights Commission, *The struggle to be seen, the power in being heard: Community insights from the Seen & Heard project* (Report, 2026) 25. See also, Kamp A, N Denson, R Sharples and R Atie, "Asian Australians' Experiences of Online Racism during the COVID-19 Pandemic." *Social Sciences* 11(5): 227 (2022); The Islamophobia Register, *Islamophobia in Australia Report V 2023-2024* (Report, 2025).
- 5 See generally, Delia Rickard, *Statutory Review of the Online Safety Act 2021* (Report, October 2024).
- 6 Delia Rickard, *Statutory Review of the Online Safety Act 2021* (Report, October 2024) 48-63.
- 7 Mario Peucker, Franka Vaughan, Jo Doley, and Tom Clark, *Understanding Reporting Barriers and Support Needs for Those Experiencing Racism in Victoria* (Report, 2024) 21-22. See also, eSafety Commissioner, *Digital use and risk: Online platform engagement among children aged 10 to 15* (Report, July 2025) 36, 40.
- 8 Without detailed legislation it remains unclear what exact risks would need to be identified and mitigated under a Digital Duty of Care. However, it is possible that this could include harms stemming from online hate speech. Delia Rickard, *Statutory Review of the Online Safety Act 2021* (Report, October 2024) 57-61.
- 9 Delia Rickard, *Statutory Review of the Online Safety Act 2021* (Report, October 2024) 58-61.
- 10 For instance, an Australian-based survey found that online racism targeting Aboriginal and Torres Strait Islander people is prevalent, with First People respondents reporting the highest likelihood of experiencing cyber racism (20.7% of First People participants). Participants of North African and Middle Eastern backgrounds and Muslim participants also reported high rates of experiencing online racism (12% and 13.2% respectively). See Andrew Jakubowicz et al, *Cyber Racism and Community Resilience: Strategies for Combating Online Race Hate* (Palgrave Macmillan Cham, 2017) 75.
- 11 For instance, Palestinian, Muslim, Arab, and Jewish participants in the Commission's Seen & Heard consultations described social media platforms as major sites of racism, harassment and hate speech. Many participants reported receiving violent threats, being doxed or targeted with racial slurs and abuse, as well as encountering online misinformation and disinformation containing racist portrayals of and harmful narratives about their communities. See Australian Human Rights Commission, *The struggle to be seen, the power in being heard: Community insights from the Seen & Heard project* (Report, 2026) 46.
- 12 Australian Electoral Commission, *Disinformation and Misinformation* (Factsheet) <https://www.aec.gov.au/About_AEC/files/eiat/eiat-disinformation-factsheet.pdf>.
- 13 Australian Human Rights Commission, *National Anti-Racism Framework Scoping Report* (Report, December 2022) 150.
- 14 Attorney-General's Department, 'Albanese Government passes legislation to combat antisemitism, hate and extremism' (Media Release, 21 January 2026). The Act does not introduce a standalone criminal offence for racial vilification. It targets racial vilification by escalating criminal penalties, expanding visa cancellation powers and targeting extremist networks.
- 15 These include section 752 of the *Criminal Code Act 2002* (ACT), sections 93ZA and 93ZB of the *Crimes Act 1900* (NSW), Chapter 7A of the *Criminal Code Act 1899* (Qld) and the *Criminal Code (Prohibited Symbols) Regulation 2024* (Qld), Part 6A of the *Summary Offences Act 1953* (SA), sections 6C and 6D of the *Police Offences Act 1935* (Tas), Division 4C of the *Summary Offences Act 1966* (Vic), Chapter 11A of the *Criminal Code Act Compilation Act 1913* (WA), and section 20A of the *Anti-Discrimination Act 1992* (NT). See also, Australian National Security, *Prohibited hate symbols* (Web page, 29 January 2026).
- 16 Australian Human Rights Commission, *National Anti-Racism Framework Scoping Report* (Report, December 2022) 152. See also, Australian Muslim Advocacy Network, Submission No 3 to the Select Committee on Social Media and Online Safety, Parliament of Australia, *Inquiry into Social Media and Online Safety* (21 December 2021) 8; and Australian Muslim Advocacy Network, Submission No 52 to the Legal Affairs and Safety Committee, Queensland Legislative Assembly, *Inquiry into Serious Vilification and Hate Crimes* (12 June 2021) 8.
- 17 Ibid.
- 18 Ibid.
- 19 Bronwyn Carlson and Tamika Worrell, *Relational Futures: Indigenous Sovereignty and the Governance of Artificial Intelligence (AI)* (Report, Global Indigenous Futures Research Centre, 2026). See also, Vishal Rana, 'Indigenous Data Sovereignty: A Catalyst for Ethical AI in Business', *Business & Society* 64(4) (22 August 2024).
- 20 The Maiam nayri Wingari Indigenous Data Sovereignty Collective defines Indigenous Data Sovereignty as the 'right of Indigenous people to exercise ownership over Indigenous Data. Ownership of data can be expressed through the creation, collection, access, analysis, interpretation, management, dissemination and reuse of Indigenous Data.' The Collective also defines Indigenous Data Governance as the 'right of Indigenous peoples to autonomously decide what, how and why Indigenous Data are collected, accessed and used. It ensures that data on or about Indigenous peoples reflects our priorities, values, cultures, worldviews and diversity.' See Maiam nayri Wingara, *Indigenous Data Sovereignty Communique Indigenous Data Sovereignty Summit 20th June 2018 Canberra, ACT* (Web page) <<https://www.maiamnayriwingara.org/mnw-principles>>.
- 21 A recent government-level example is the co-design and adoption of the [Framework for Governance of Indigenous Data](#) in the Australian Public Service.
- 22 Examples include restorative justice programs and transformative justice and community accountability measures that aim to repair harm, address drivers of racism before they manifest into violent speech or conduct and holistically tackle both interpersonal and systemic racism. These alternatives are part of broader efforts to move the criminal justice system away from punitive solutions and incarceration, which disproportionately target Aboriginal and Torres Strait Islander peoples, as well as some other negatively racialised communities. See also, Australian Human Rights Commission, *National Anti-Racism Framework Scoping Report* (Report, December 2022) 152, 154, 161.
- 23 Australian Human Rights Commission, *National Anti-Racism Framework Scoping Report* (Report, December 2022) 151-153.