Online Hate Speech In Australia

The Role of News Media and Pathways for Change

PURPOSE

We acknowledge the Traditional Custodians, in particular Gadigal, Wangal and Tharawal peoples, and Wurundjeri of the Kulin Nation, upon whose unceded ancestral lands we live and work. We pay respect to their Elders, past and present, and acknowledge the pivotal role that Aboriginal and Torres Strait Islander people continue to play within the Australian community.

About Purpose

Purpose is a global social impact agency that works towards a more open, just and habitable world.

We come to this research with a strong perspective that change is needed to reduce harm and include more people to thrive in our democracies.

Our research work is oriented towards supporting actions at all levels, from individual action, community programs and policy change.

Acknowledgments

We would like to acknowledge the support and advice from a number of organisations who were stakeholders in this project:

The Archewell Foundation, whose vision and support for this work to identify and explore difficult problems facing our democracies is an important and impactful contribution.

Demos, for exchanging ideas and feedback through our concurrent research, and for their major contribution through their published works in this area.

Rita Jabri Markwell and Associate Professor Sukhmani Khorana who participated in highly valuable exploratory interviews on this research.

Our earlier work with Reset Australia and the Sydney Policy Lab, led by Dhaksh Sooriyakumaran and Jananie Janarthana, underpinned our approach and perspective on this research.

Our use of a symbol system to provide readers with an informed way to interact with traumatic material is drawn from Building Consentful Tech by Una Lee, Dann Toliver and their colleagues.

Lastly, this project would not have been possible without the efforts of a dedicated team across Purpose. Researching online hate and disinformation can be personally exhausting and demoralising. We would like to acknowledge this difficult and important work to establish and deliver research about online threats.

Contents

📍 **Please click each link to view**

Executive Summary

Online hate speech is a particularly visible and toxic trait of social division. It fosters the negative targeting of communities, diminishes peoples' health and wellbeing, erodes social cohesion and also negatively influences the design of Government policies and programs.

In Part Two of this report we outline promising pathways to addressing the roles that powerful news media institutions play in harmful online activity, particularly dehumanising hate speech. While previous quantitative and qualitative research has helped to highlight the problem, there is a lack of contemporary rich case studies that illuminate the range of stakeholders, narratives and tactics around online hate in Australia.

In Part One of this report, we articulate the patterns of 'deficit discourse' in news media publishing (content that negatively frames and scapegoats a particular group of people), and the mutualistic relationships between news media actors and external actors who are producing and amplifying strong deficit discourse or even engaging in hate speech. These patterns were illustrated through four case studies that also challenge fixed notions of responsibility and highlight practical dilemmas for news media practitioners and other stakeholders.

Here in Part Two, we explore these challenges by reviewing the regulatory landscape in Australia and possible pathways for improvements. We begin by describing the patchwork of Federal and State legislation and regulation that applies to online hate speech and news media publishing in Australia. We then explore the challenges and loopholes posed by this approach, and we examine possible solutions, and barriers, to address these issues. Finally, we explore other opportunities, outside the regulatory framework, that could reduce the prevalence and harm of online hate. These approaches include tools for greater moderation, monitoring and analysis of dehumanising hate speech, to education initiatives for journalists, as well as coalitions for greater research and knowledge sharing.

We conclude the report with recommendations to address these challenges and continue building our understandings, summarised here:

1. Further Research

The issue of dehumanising hate speech in Australia, and elsewhere, needs further investigation, which will strengthen the evidence base for policy and program design. Such research should cover different patterns of online hate and their impacts on specific communities; specific tactics or spaces that host online hate activity; and the roles played by different actors within our information ecosystem.

2. Protect freedom of expression while strengthening hate speech laws

Effective hate speech legislation is the backbone of a holistic approach to countering dehumanising hate speech. The limitations of existing laws are demonstrated clearly in the case studies of Part One of this report. Priorities for improvement include:

- national definitions for dehumanising speech and discourse;

- broadening the definition of groups covered by relevant laws;

- penalties for hate speech and disinformation within the online safety framework – including civil penalties;

- clarifying that hate speech and discrimination laws apply to social media companies based overseas; and

- explicit rules and clarity on the balance between freedom of expression and other fundamental rights.

3. Address patchwork of regulation on news media

We suggest expanding and updating the powers of the media regulator to deal appropriately with the complexity of modern online publishing. Industry standards need to be drafted by regulators – not industry – in better consultation with communities and experts, and with adequate resourcing for enforcement.

4. Broaden diversity of news media ownership

The long-running monopolisation of the news media sector has led to a lack of accountability, and an outsized level of influence, in dictating public narratives, and in shaping political debate and industry practice. Governments, institutions and philanthropy need to dramatically increase their support for a broader diversity of news media publishers. This will require stronger media diversity and antitrust legislation as well as Government incentives, and philanthropic investment, in locally accountable and community-based news outlets to provide healthier competition.

5. Impose greater accountability on social media companies

The role of social media companies in enabling and amplifying dehumanising hate speech is well established and regulation to hold them accountable is overdue. Priorities here should include enforced transparency and algorithmic audits by regulators and researchers, accountability for amplifying harmful content and comments, and the establishment of a statutory duty of care.

6. Greater investment in media literacy programs

Responsibility for detecting and countering hate speech should not sit with the public, but there is a need to improve communities' and individuals' media and digital literacy - as this enables their full participation in society. The scale of this requires investment by governments and industry, and should include building awareness of relevant legislation and industry codes of practice. This will better equip communities and individuals to understand, identify and report dehumanising hate speech – as well as knowing their rights in relation to the issue.

7. Support and expand educational programs for journalists and news media practitioners

These should cover the impact that their work has when it reproduces or provokes the amplification of deficit discourse (see Key Definitions), dehumanisation and hate speech.

Key Definitions

Key Definitions

Deficit Discourse

Deficit discourse is a durable system of language, story and meaning that positions certain groups as lacking, incompetent, a threat or a problem to be solved (L. P. M. Davis & Museus, 2019; Fforde et al., 2013). Hate speech and dehumanising speech are often a visible and extreme end of a more subtle and pervasive deficit discourse. In this report we note the presence of deficit discourse targeted at a number of different groups, including First Nations people, Indian migrants, transgender people and people who came to Australia as refugees.

In their study of "how narrative framings of Aboriginal and Torres Strait Islander people are reproduced in policy", Fogarty et al. provide this description of how deficit discourse works:

> In certain discursive spaces in Australia, the term "Aboriginal and Torres Strait Islander people" has come to be associated with particular negative tropes, such as being unhealthy, undereducated, unemployed, violent and socially dysfunctional. […] Such tropes of deficiency reduce and homogenise people, and tell us nothing about their complex lives and socio-economic circumstances. (Fogarty, Bulloch, et al., 2018, pp. 2–3) … It is crucial to note that in analysing and mapping discourses of deficit, we do not seek to "problem deflate". Aboriginal and Torres Strait Islander Australians face well-documented realities of socio-economic "disadvantage". Discourses of deficit occur when discussions and policy aimed at alleviating disadvantage become so mired in reductionist narratives of failure and inferiority that Aboriginal and Torres Strait Islander people themselves are seen as the problem. These discussions thus become a continuation of pejorative and patronising race-based discourses in terms of which Aboriginal and Torres Strait Islander people have long been represented (Fogarty, Bulloch, et al., 2018, p. 30)."

Fogarty et al. note that the "relationship between politicians and the media is also writ large in the deficit discourses reproduced in the media", and that

these discourses have contributed to shaping health policies and outcomes (Fogarty, Bulloch, et al., 2018, p. 5).

In their review of scholarship on "deficit thinking", Lori Patton Davis and Samuel D. Museus highlight a number of related terms, often used interchangeably, including: deficit thinking, -discourse, -paradigm, -ideology and -assumptions (L. P. M. Davis & Museus, 2019). For this study we have chosen to use "deficit discourse" following publications by the Lowitja Institute and others in the area of First Nations' health in Australia, which are particularly relevant to our findings (Fforde et al., 2013; Fogarty, Bulloch, et al., 2018; Fogarty, Lovell, et al., 2018).

"Dehumanising Speech" and "Dehumanising Hate Speech"

Dehumanising speech is a related concept to hate speech that outlines some of the particularly harmful tactics used to attack certain groups (Maynard & Benesch, 2016, p. 80). Broadly, it seeks to justify the unequal treatment of targeted peoples by presenting them as lacking humanity.

We consider dehumanising speech to be a useful and descriptive term for what a certain kind of hate speech does. Dehumanising hate speech frequently operates as the extreme and visible edge of a broader, deficit discourse.

> Dehumanising hate speech frequently operates as the extreme and visible edge of a broader, deficit discourse.

Because of the nuance and difficulty involved in identifying hate speech at scale, the Australian Muslim Advocacy Network (AMAN) has argued that focusing on dehumanising speech can "cut through these challenges and focus on some of the most potent vectors of harm", based on their research into anti-Muslim content and other forms of online hate. AMAN has published a precise definition of dehumanisation, which is provided in full under Appendix A (Australian Muslim Advocacy Network, 2023).

Hate Speech

"Hate speech" most commonly refers to expressive acts that incite discrimination, hostility or violence towards a person or group because of who they are, or on the basis of their protected characteristics. There is no standardised definition or list of protected characteristics applied across contexts and jurisdictions. In Australia, there are relevant definitions of vilification in federal and state legislation. However, these definitions vary, as do the lists of protected attributes. This patchwork of definitions has been identified as a significant problem in our Part Two report.

Additionally, the term "hate speech" is often used in the heat of arguments about what people are publishing, what motivates them, and how speech should be regulated, and so the definition of the term itself becomes contested (Gagliardone et al., 2014). Mindful of the importance of clear definitions within research and policy, we have chosen to use social media company Meta's definition of hate speech. This highlights the varied and nuanced nature of hate speech as an attack, and has particular relevance to our research context. Meta defines hate speech as:

> **"** a direct attack against people – rather than concepts or institutions – on the basis of what we call protected characteristics: race, ethnicity, national origin, disability, religious affiliation, caste, sexual orientation, sex, gender identity and serious disease. We define attacks as violent or dehumanising speech, harmful stereotypes, statements of inferiority, expressions of contempt, disgust or dismissal, cursing and calls for exclusion or segregation (Meta, 2023)."

A list of additional resources and definitions of hate speech, particularly from online platforms, is provided under Appendix B.

News Media

"News media" refers to a diverse collection of organisations and individual practitioners who publish news reporting, opinion and debate about current affairs. Traditionally, news media includes printed newspapers and magazines, radio, and broadcast television. As the internet penetration rate continues to increase, news media further encompasses other formats, such as online news articles, online news video, online radio, blogs, podcasts and social media-based reporting.

We refer readers to a recent online review of evidence about Australia's media ownership and news audiences provided by RMIT ABC Fact Check (RMIT ABC Fact Check, 2021). It presents a large collection of statistics on the demographics and audience sizes of Australian news media across a range of formats.

For those unfamiliar with Australia's news media landscape, popular outlets are listed in Appendix C.

Social Media

In this report, we use "social media" to refer to the range of social media platforms listed below and also to the broad phenomenon of online social activity on them.

A non-exhaustive list of platforms includes: Facebook, YouTube, Instagram, WhatsApp, Twitter, LinkedIn, Snapchat, Pinterest, TikTok, Reddit, Telegram, WeChat, BitChute, Gab, Rumble and Odysee.

'Strength-based' Approaches

Strengths-based approaches is a term that is used differently across the literature. In this report, our use of this term is informed by the work of the Lowitja Institute, where they consider it to be a number of grouped approaches including 'asset-based approaches, resilience, cultural appropriateness, social determinants of health and ecological theories, protective factors, empowerment, holistic approaches, wellness and wellbeing, strengths-based counselling approaches and positive psychology, decolonisation methodology, and salutogenesis' (Fogarty et al, 2018).

Acronyms

ABC	Australian Broadcasting Corporation (public broadcaster)
ACMA	Australian Communications and Media Authority
AFN	Australian Financial News
AMAN	Australian Muslim Advocacy Network
AMLA	Australian Media Literacy Alliance
LGBTQIA+	Lesbian, Gay, Bisexual, Transgender, Queer, Intersex, and Asexual
MEAA	Media, Entertainment & Arts Alliance
NITV	National Indigenous Television (public broadcaster)
NSW	New South Wales (Australian State)
RMIT	Royal Melbourne Institute of Technology (university)
SBS	Special Broadcasting Service (public broadcaster)

Introduction

Introduction

The case studies presented in Part One of this report illustrate the troubling dynamics of deficit discourse and mutualism in the relationships between news media and other actors. Part Two discusses the ineffectiveness of regulatory frameworks and approaches in addressing these patterns. First, we discuss some key improvements to regulations that could address hate speech online. We then suggest possible regulation that takes into account the broader dynamics of dehumanising hate speech and modern news media publishing in Australia.

Current federal anti-discrimination legislation is designed to protect individuals and communities from hate speech, and is complemented by varying forms of regulation of news media and online platforms. However, this approach is increasingly outdated and ill equipped for the modern digital environment. Our research demonstrates the need for an expanded view of the issue, and raises key questions about the:

- responsibility of journalists and newsrooms;

- incentives embedded in the business models of news and social media companies;

- role of an engaged and informed public; and

- impact of a news media industry lacking in diversity of ownership.

Given the complexity of our modern information ecosystems and the ways that dehumanising hate speech manifests online, there is no legislative silver bullet available. However, our recommendations represent a significant shift in how hate speech is considered, and focuses on the underlying elements that contribute to its proliferation. This report recommends a more

holistic 'strengths-based' approach that includes both regulatory and non-regulatory mechanisms; some of which will require measured public debate and education initiatives to refine.

While better regulation of news and social media companies is necessary, we do not suggest that these companies have created the issue of hate speech nor that they are solely responsible for its proliferation in Australia. However, the case studies make it clear that these companies intentionally, or inadvertently, amplify, enflame and worsen the existing cultural and historical dynamics of hate speech. Effective regulation is, and needs to remain, a primary goal but, without concerted efforts to shift the awareness, resilience and relationships communities and civil society have with news and online platforms, we will see online hate speech worsen. Our media, information ecosystems, power structures and business models are littered with incentives for actors to directly and indirectly foster the proliferation of dehumanising hate speech online.

Without concerted efforts to shift the awareness, resilience and relationships communities and civil society have with news and online platforms, we will see online hate speech worsen.

Our Approach

This two-part research report is designed to support those working to develop policies and programs that respond to the proliferation of dehumanising hate speech online. This second part focuses on existing regulations, and the debates on policies and other, potentially useful, ways of mitigating the issues outlined in Part One's case studies.

This part includes a survey of relevant Government policy and legislation, along with a discussion of how these measures apply to the patterns of activity explored in Part One. We identify loopholes and challenges for regulation and discuss the barriers to policy change, as well as briefly exploring promising opportunities outside Government policy and regulation.

Our approach here included conducting interviews with experts on these policies and issues, along with desk research that included:

- a review of existing legislation;

- outlining the relevance of existing legislation with the case studies in Part One of this report; and

- a literature review of how different groups and researchers are working to curb hate speech.

1. The Existing Regulatory Landscape

The Existing Regulatory Landscape

This section outlines the patchwork (and gaps) of relevant regulation in Australia. For ease of communication, we use short forms when referring to broad stakeholder groups, with the knowledge that not all entities or actors within those groups play an equal role in causing or moderating hate speech online. Those stakeholder groups include:

Stakeholder Groups	Categories
News media companies	Broadcast, print, radio, online, corporate, public
Social media platforms	Foreign-owned, major platforms, 'alt' platforms
Regulators	Consumer, news media and communications, online safety, industry bodies
Community	Civil society groups, community or social groups, individuals

The complex nature of how dehumanising hate speech germinates and proliferates online requires a combination of stakeholders and accountability mechanisms. Within Australia – as in much of the rest of the world – the patchwork of regulation and frameworks is intended to collectively address the issue. The regulations discussed in this section loosely fall within three overlapping categories:

1.1 Hate Speech, Discrimination and Vilification	1.2 News Media and Journalists	1.3 Social Media & Digital Platforms
Federal and State legislation	Radio & TV industry codes of practice; Australian Press Council standards of practice	Online Safety Legislation; company-specific guidelines; legal precedents for example, the Voller case (High Court of Australia, 2021)

1.1 Hate Speech, Discrimination and Vilification

Federal and state legislation on hate speech, discrimination and vilification is intended to serve as the legal foundation for the prevention and redress of such behaviour across all industries and sectors in Australia. Each federal, state, and territory jurisdiction has adopted a different approach (see Figure 1). In addition, there is a mix of civil and criminal regulations that vary across jurisdictions. There is no single legal definition of hate speech within Australia, no uniform approach to vilification, and hate crime laws do not provide the same degree of protection or coverage for groups based on their protected characteristics (see Key Definitions). For individuals and communities this means that opportunities for redress or justice can vary depending on their identity or background as well as the state or territory in which they live.

At federal level, the legislated threshold is committing an act that is reasonably likely to offend, insult, humiliate or intimidate another person or group of people because of their race, colour or national or ethnic origin. The threshold under State-level legislation is inciting hatred towards, serious contempt for, or severe ridicule of, a person or group of people on the grounds of their race.

The legislation on hate speech of a racial nature is by far the most comprehensive in Australia, and covers various expressive acts including speaking, gesturing, drawing and written publications. There are three essential components of such conduct.

1. The words, sounds, images or written communication are expressed publicly.

2. The act must be reasonably likely to offend, insult, humiliate or intimidate the people against whom it is directed.

3. It must be done because of the race, colour or national or ethnic origin of the group against whom it is directed.

Exemptions to these rules include acts done 'reasonably and in good faith for academic, artistic, scientific or research purposes'; or acts in the public interest (including discussion or debate). Importantly, the burden of proof is on the complainant. For instance, the complainant must prove that the act:

- was carried out in public;

- was carried out because of their race; and

- would be reasonably likely for someone of the same race to be offended.

The respondent must establish that the act is covered by one of the exceptions.

People from a specific group who have been targeted may initially seek redress through a conciliation process run by the Human Rights and Equal Opportunity Commission. If the complaint cannot be resolved, or is found not to have merit (or has been dealt with by another agency), it will be terminated. In this case, the only option for redress for an aggrieved
person is to take their complaint to the Federal Circuit Court or the Federal Court of Australia.

While there are several pieces of related federal legislation, the foundational components are under Sections 18C and 18D of the Racial Discrimination Act 1975 (AustLii, n.d.). The Keating Government amended the Act to introduce these sections in the mid-1990s after inquiries, including the National Inquiry into Racist Violence in Australia (1991) and the Royal Commission into Aboriginal Deaths in Custody (Australasian Legal Information Institute, 1998), revealed that targets of racial violence had little recourse to civil remedies under common law. The amendments introduced these sections to address gaps in the law, and provided a legal framework that recognised the harm caused by racial harassment and abuse.

A significant number of cases proceed from the conciliation process to legal action, such as Eatock v Bolt in 2011 (ATNS Project, 2020). Here, a commentator for News Corp publications, Andrew Bolt, was found to have breached Section 18C by publishing racially offensive comments about light-skinned Aboriginal people. The Federal Court of Australia found that Bolt's comments were likely to offend, insult, humiliate or intimidate individuals of mixed race and were not made in good faith. As a result, Bolt was ordered to remove the offending articles from his website, publish a correction and apology, and refrain from publishing similar comments.

While this ruling was a significant victory for those who advocate for stronger protections against hate speech, extensive resources were required to move the case through the Federal Court – an avenue unavailable to many Australians

> **Sections 18C and 18D of the Racial Discrimination Act 1975**
>
> Section 18C makes it unlawful to commit a public act that is 'reasonably likely, in all the circumstances, to offend, insult, humiliate or intimidate a person' (or group of people) 'because of their race, colour or national or ethnic origin'. The standard to be met in order for this section to be contravened is conduct which has profound and serious effects, not to be likened to mere slights. However, Section 18D provides exemptions for artistic works, academic or scientific inquiry, fair and accurate reports of matters in the public interest, and fair comments of matters in the public interest which express the genuine belief of the person making the comment.

– and the effectiveness of the ruling in deterring racial vilification has been questioned (Clark, 2013). Others argue that the decision may have had a chilling effect on freedom of speech and that it could be used to stifle legitimate debate and criticism (Alcorn, 2014). This particular criticism plays a role throughout conversations of news and social media regulation, but is often used as a political tool that elevates the right to freedom of expression as an absolute, even when used to undermine a person's fundamental right to dignity.

Both approaches focus on the *impact* of the act on the audience and whether it is likely to incite hatred in others, an important dynamic here is that there is no need to prove the intention to incite, rather the focus is on the effects of the public act. While this approach has its strengths, it does require action to be taken by affected individuals, which assumes that such means is accessible by all.

Matrix of hate speech protections and legislation at State and Federal levels in Australia

	COMMONWEALTH (FEDERAL)	ACT	NSW	NT	QLD	SA	TAS	VIC	WA
Legislation	• Racial Discrimination Act 1975 • Sex Discrimination Act 1984 • Human Rights Commission Act 1986 • Disability Discrimination Act 1992 • Age Discrimination Act 2004 • Fair Work Act 2009 • Sex Discrimination and Fair Work Amendment Act 2021	Discrimination Act 1991 (ACT)	Anti-Discrimination Act 1977 (NSW)	Anti-Discrimination Act 1996 (NT)	Anti-Discrimination Act 1991 (QLD)	• Civil Liability Act 1936 (SA) • Equal Opportunity Act 1984 (SA)	Anti-Discrimination Act 1998 (TAS)	• Racial and Religious Tolerance Act 2001 (VIC) • Equal Opportunity Act 2010 (VIC)	Equal Opportunity Act 1984 (WA)
Inclusions									
Race	✔	✔	✔	✔	✔	✔	✔	✔	Under Federal
Sex	✔	Under Federal	Under Federal	Under Federal	Under Federal	Under Federal	✔	Under Federal	Under Federal
Age	✔	Under Federal	Under Federal	Under Federal	Under Federal	Under Federal	Under Federal	Under Federal	Under Federal
Disability	✔	✔	Under Federal	Under Federal	Under Federal	Under Federal	✔	Under Federal	Under Federal
HIV/Aids	✔	✔	✔	✗	✗	✗	✗	✗	✗
Sexuality	✗	✔	✔	✗	✔	✗	✔	✗	✗
Gender Identity	✗	✔	✔	✗	✔	✗	✗	✗	✗
Intersex	✗	✔	✗	✗	✗	✗	✔	✗	✗
Religion	✗	✔	✔	✗	✔	✗	✔	✔	✗

1.2 News Media: Regulation, Codes, Standards

The regulation of news media in Australia is primarily the responsibility of the Australian Communications and Media Authority (ACMA) and the Australian Press Council. ACMA is an independent statutory authority of the Commonwealth which sets and oversees the rules about communications and media services and markets, and addresses complaints about alleged breaches. The Australian Press Council is responsible for setting good standards of media and journalism practice for its members, and promoting freedom of speech. The Press Council is the principal body for responding to complaints about newspapers, magazines and online-only publications. Member publications, journalists and associations agree to cooperate with the council's investigations into complaints and to publish subsequent adjudications.

Radio & Television Broadcasters | ACMA Industry Codes of Practice
ACMA has worked with free-to-air television and radio broadcasters to adopt several self-regulatory codes of practice covering their commercial content and to help viewers make informed choices about their viewing (ACMA, 2023). These codes have been designed in accordance with a vaguely defined concept of 'current community standards' and aim to ensure that content 'not suitable' for broadcast (i.e., offensive, inappropriate, disturbing or harmful to certain groups of people) is not aired.

The Commercial Radio Code of Practice (Commercial Radio Australia, 2017) and *Commercial Television Industry Code of Practice* (Free TV Australia, 2015) define material not suitable for broadcast in a similar manner; as that likely to incite in a 'reasonable' person, 'hatred' or 'intense dislike' of any person or group of persons because of age, ethnicity, nationality, race, gender, sexual preferences, religion or disability. The radio code also covers 'transgender status' in its named protected characteristics, while the television code adds 'colour'. These codes also acknowledge that there may be exceptions for material presented reasonably and in good faith, such as for academic, artistic, religious instruction, or scientific purposes, or when discussing a matter of public interest.

In 2018, the Channel Seven morning show Sunrise faced criticism, and was found to have breached the commercial television code of practice, for a segment discussing the removal of Indigenous Australian children

> These codes have been designed in accordance with a vaguely defined concept of 'current community standards'.

from their families. The ACMA's investigation (BI-363 Sunrise) outlines their assessment and concluded that:

" the inclusion of inaccurate information and the sweeping statements such as 'they're getting abused, they're getting hurt and they're getting damaged' and the presenter's concluding comment 'poor kids', presented the issues in a polarising and unreasonable manner which did not demonstrate good faith. Comments about the threat to children in Indigenous communities were not broadcast within any framework that would enable viewers to judge whether there was a reasonable objective basis for the attitudes displayed. The licensee created a scenario in which negative associations accrued around Indigenous people because of the framing of the issue which:

- emphasised and repeated in connection with Indigenous communities that they were dangerous environments for children

- unfairly contrasted 'safe' white families with 'dangerous' Indigenous families; and

- suggested that the forced removal of Aboriginal children from their families has been beneficial in the past and should be pursued again.

As a consequence, Indigenous people were presented as being inferior to non-Indigenous people in their ability to raise their children in a safe environment.
(ACMA, 2018)

ACMA ordered Channel Seven to apologise for this and to provide training for staff. Channel Seven sought to challenge the finding in court while, separately, Aboriginal elders and young leaders, including Aunty Rhonda Dixon-Grovenor and Aunty Debra Swan, lodged a complaint with the Human Rights Commission, and Yolngu woman Kathy Mununggurr and 14 members of the Yirrkala community sued for defamation for the use of their images in the segment. That case was settled out of court, and the eventual controversial apology occurred almost two years after the segment was broadcast (Dye, 2019).

In another instance in 2019, commercial radio station 2GB was found to have breached ACMA's code of practice during a segment hosted by Alan Jones who made the following suggestions for how the then-Prime Minister Scott

Morrison should respond to New Zealand Prime Minister Jacinda Ardern's criticism of particular Australian policies (emphasis included in the ACMA's investigation report):

> **"** 1) **Now I hope Scott Morrison gets tough here with a few backhanders**…he's got a stack of staff, someone should have him, this morning, with a full briefing. [2.55]
>
>   2) Scott Morrison I repeat, has got a stack of staff, I hope someone's given him a full briefing, and I **hope he goes for the throat this morning.** [8.18]
>
>   3) **I just wonder whether Scott Morrison's going to be fully briefed to shove a sock down her throat.** I mean she is a joke this woman. An absolute and utter lightweight. [31:15]

(ACMA, 2020)

ACMA found the broadcaster had violated the code's accepted standards of decency and accuracy, but was not in breach of the code in relation to the encouragement of violence or brutality to vilify people on the basis of gender. In response, Jones apologised on air and 2GB said it would incorporate the investigation›s findings into staff training.

In both matters, ACMA conducted investigations and found the media outlets had breached the code. This shows that ACMA can be somewhat effective in regulating hate speech, enforcing industry codes of practice and encouraging media outlets to take steps to prevent similar breaches. It is worth noting, however, that some critics – such as journalist Jonathan Holmes – have argued ACMA is limited in its ability to investigate and to impose strong enough sanctions to deter major broadcasters (Burrowes, 2010). There is a debate over whether ACMA should be able to impose more significant deterrents, and there are lingering concerns over the fact that ACMA does not have power to regulate hate speech on online platforms or social media.

Print and Online News Media Outlets | Australian Press Council Standards of Practice, Principles & Guidelines
As for the regulation of print and online media outlets, the Australian Press Council has adopted several standards to regulate commercial publications and to provide readers with ways of voicing their complaints and concerns. The Australian Press Council says the standards are intended to uphold the freedom of the press while also ensuring that journalists and media outlets

adhere to strict principles of accuracy, confidentiality, fairness, privacy, integrity and transparency.

These standards prohibit newspapers, magazines, news websites, and other types of online publications (such as content produced by individual journalists and bloggers) from publishing material which is likely to be significantly inaccurate, misleading, indistinguishable from other material (including opinion), unbalanced, unfair or adverse. However, the standards also acknowledge there may be exceptions for material presented reasonably and in good faith, such as for academic, artistic, religious instruction, or scientific purposes, or when discussing a matter of public interest in a fair comment.

In 2021, the Australian Press Council received a complaint about an online article published by the Daily Telegraph headed 'Transgender Sport Safety and Fairness Concerns Raised by Female Volleyball Players'. The article raised concerns about transgender athletes playing in women's competitions, and included comments such as 'it would be fairer if [transgender players] played in existing mixed game competitions [rather than female competitions]', and that '[women] are now being asked to sacrifice our own competitions to boys and men who simply want to play in [female] sports codes'. The primary concern was that there was a fundamental lack of balancing comments from a transgender athlete, a LGBTQIA+ sporting association, or any individual or organisation in support of transgender participation in female sporting competitions. In response, the Australian Press Council conducted an investigation and found that the publication failed to take reasonable steps to ensure that factual material was presented with reasonable fairness and balance (and, subsequently, was in breach).

In another instance in 2020, the Australian Press Council received a complaint about an article published by the Sunday Telegraph in print headed 'Where's The Real Justice?'. The article commented on the Black Lives Matters protests in Australia during June 2020, and asserted that 'the reality in this country - and in the US - is that the greatest danger to Aboriginals and negroes is themselves', and that 'blaming police and the corrective services system for their ills is, frankly, a cop-out'. Here, the primary concern was that the terms contained within the article were outdated racial slurs, and that the article implied that Indigenous Australians and African Americans were not only responsible for, but deserving of, such violence. In response, the Australian Press Council conducted an investigation and found that the publication failed to take reasonable steps to avoid substantial offence, distress or prejudice (without sufficient justification in the public interest), and failed to ensure the writer's expression

of opinion was not based on significantly inaccurate material or omission of key facts (and, subsequently, was in breach of General Principle 3 and 6).

These cases demonstrate the Council's lack of effectiveness in enforcing industry standards of practice as, despite finding both outlets in breach of them, the Press Council does not have powers to enforce sanctions on their members. Former Prime Minister Kevin Rudd and the Media, Entertainment and Art Alliance (MEAA), among others, have criticised this lack of enforcement powers (and lack of independence) with Mr Rudd

The MEAA, which represents 5,000 journalists, quit the Australian Press Council in 2021.

calling the Council a 'toothless tiger' (2021). The MEAA, which represents 5,000 journalists, quit the Australian Press Council in 2021 stating that, 'despite media convergence being a lived reality for journalists and the public for a decade, the regulatory framework had failed to keep up to date' (MEAA, 2021). As a self-regulatory body, there is no requirement for publishers to join, or adhere to, the standards of the Australian Press Council. The problems with relying on the Council for redress is further exemplified by the explicit dehumanising hate speech from Australian Financial News (AFN) outlined in Case Study 2 of this report, as AFN is not a member of the Council, rendering its codes of practice irrelevant.

1.3 Social Media and Online Platform Regulation

There are few policies on the regulation of digital platforms on hate speech, with a heavy reliance on weak self-regulatory approaches that tend to lack accountability, transparency and local contextualisation. What legislation there is, in Australia and elsewhere, tends to have a narrow focus on harms affecting young people and children. Typically, most regulation falls under the community guidelines developed and enforced by digital platforms themselves.

eSafety Commissioner

The Australian eSafety Commissioner is a world-first regulator with a remit of keeping Australians safe online with powers to address cyberbullying, image-based abuse and illegal and harmful content. The legislative functions of the Commissioner include:

- powers to compel online platforms to remove harmful content;

- setting 'Basic Online Safety Expectations' to ensure social media, messaging, gaming and other digital apps take reasonable steps to develop safe products; and

- the power to require reporting for transparency and accountability purposes.

The Commissioner also provides avenues for complaints, has investigative and removal powers, and has several enforcement options including being able to seek civil penalties for a failure to comply.

While the protection of children's safety online is a priority for the eSafety Commissioner, enhanced powers were also granted in 2021 to protect Australian adults from cyber abuse. The abuse needs to meet two key criteria of:

1. intention to cause serious harm; and

2. being menacing, harassing or offensive.

The Commissioner assesses whether a reported act targets an individual based on 'their cultural background, gender, sexual orientation, disability, mental health condition or family or domestic violence situation'.

This regulatory scheme includes specific exemptions to protect and balance freedom of expression with the need to protect individuals from abuse. However, critics claim these powers constitute significant overreach; limiting freedom of expression and representing a slippery slope towards government censorship (Wilson, 2021).

The eSafety Commission is designed as a backstop when online platforms' own complaint mechanisms fail to meet the needs of the victim and, in this way, does not conduct proactive investigations into harmful content online beyond child abuse material.

Platform-specific Community Guidelines

All digital platforms have community guidelines of some form which are used by their teams and systems to enforce moderation of content and user behaviour. The sheer scale of content that requires moderation on these platforms poses a challenge for companies in assessing what content is acceptable, what needs to be removed and how to do this efficiently (Jiang et al. 2020) analysed the community guidelines of the 11 largest social media platforms and identified 66 types of rules across major social media platforms, with more than a dozen related directly or indirectly to hate speech. Importantly, these guidelines are constantly revised in response to emerging issues. While this is necessary, the discrepancies in what forms of hate speech platforms decide to moderate (or not) have come to serve

as a competitive distinction. This is particularly apparent in relation to some online communities' perspective that moderation on mainstream platforms is censorship, enabling the rise of 'alt' platforms where hate speech content is allowed to proliferate without moderation.

As these platforms have developed their guidelines without consulting Australian communities or regulators, they also lack any external ability to enforce them. Researchers and journalists have raised concerns about the companies' inconsistent implementation of these guidelines. In 2021 the Wall Street Journal published the 'Facebook Files' which exposed the company's secretive program 'cross check' which effectively allowed the prominent accounts of 'powerful actors' to post content 'including harassment and incitement to violence' with impunity (Wall Street Journal, 2021). It has been repeatedly found that the company failed to address numerous internal complaints about the significant under-resourcing of their Safety and Integrity team, leaving them unable to meet the moderation needs of the platform's guidelines (Lomas, 2021).

Social media platforms are not accountable for the content posted by users on their platforms, including the comments section on pages and groups. Our research has observed a pattern where a post within a Facebook group or page can evade platform moderation by refraining from using explicit hate speech language; instead using deficit discourse and enabling the comments section to become an unmoderated vessel of hate speech and racism.

Relevance of the Voller Decision

This issue around accountability of hate speech in comments sections came to a head when Dylan Voller, a former Northern Territory youth detainee, was featured on the Australian Broadcasting Corporation (ABC) television program Four Corners in July 2016. The program sought to expose the mistreatment and abuse of youth in child protection and detention systems in the Northern Territory, and included footage of Voller shackled to a restraining chair and placed in a spit hood. This footage prompted then-Prime Minister Malcolm Turnbull to announce the Royal Commission into the Protection and Detention of Children in the Northern Territory.

Following the program (between July 2016 and June 2017), there was an avalanche of media stories that could serve, in its own right, as a research case study into deficit discourse and mutualism. The stories featured numerous defamatory comments posted in response to articles published on the Facebook pages of the Sydney Morning Herald, The Australian, Sky News, The Bolt Report and the Centralian Advocate. Comments included racist and dehumanising hate speech, with some publishable examples

suggesting that Voller was a 'little grub' and a 'vicious little criminal'. As a consequence, Voller brought three separate defamation proceedings against the companies that own those titles: Fairfax Media, the Australian News Channel and Nationwide News.

In 2021, the [High Court](#) upheld a finding that, in respect of defamation liability, appellant media companies are publishers of comments posted to their public Facebook pages by third-party users. The decision confirms that an organisation or individual which maintains Facebook pages or other online forums (through which they allow or encourage public interaction) might be liable for any defamatory comments made by others in response. This decision has significant implications for media companies and their use of social media or online platforms, as there is now significant legal risk and potential liability for failing to manage and moderate comments on their social pages. In response to this finding, Facebook allowed publishers and other page owners the ability to switch off comments or to select the types of accounts that are allowed to comment.

The High Court decision, as well as a failed attempt by the Morrison Government to legislatively undermine it, highlights the lack of accountability for any party in this scenario, and demonstrates that individuals are left with the burden of accessing legal representation – and the requisite resources – to achieve redress.

2. Policy Challenges and Loopholes

Policy Challenges and Loopholes

The ineffectiveness of Australia's regulatory landscape in addressing dehumanising hate speech online is due to several factors that are also observed in other countries. This section outlines some of the gaps in the regulatory framework, particularly the limited regulations covering deficit narratives and the almost complete lack of regulations in addressing mutualism.

In addition to the policy challenges outlined below there are also several contextual and systemic issues. The speed of technological developments significantly outpaces the ability of governments to update existing, or to pass new, legislation. It is worth noting that industry standards are more agile and could be updated more regularly without waiting for legislative action.

Numerous experts argue for a change in direction to policy, such as establishing a duty of care for platforms (Gelber, 2021), to address gaps and to better future-proof regulation. Passing substantive regulations is becoming increasingly difficult in many democratic societies as, ironically, the bias in news media along with attention-optimising digital platforms is increasingly polarising communities.

This section outlines some of the key contributing factors to the ineffectiveness of our regulatory framework in addressing deficit narratives, mutualism, and dehumanising hate speech more broadly:

2.1 Defining Dehumanising Hate Speech and Balancing Freedom of Expression

2.2 Inconsistencies Between News and Social Media Regulation

2.3 Enforcement Limitations

2.4 Identification, Monitoring and Removal of Hate Speech Content

2.1 Defining Dehumanising Hate Speech and Balancing Freedom of Expression

In its submission to the Parliamentary Joint Committee on Human Rights Inquiry Into Freedom of Speech (2017), the Australian Human Rights Commission noted that one of the main challenges with regulation here is defining what constitutes hate speech. The lack of a clear and comprehensive definition makes it difficult to effectively identify and address instances of hate speech, particularly in relation to deficit narratives. This has led to inconsistencies in the application of hate speech laws and a lack of clarity for individuals, news media organisations and tech companies.

Section 18C of the Racial Discrimination Act, for example, applies only to conduct that is 'reasonably likely to offend, insult, humiliate or intimidate a person or group of people on the basis of their race, ethnicity or nationality'. Organisations and individuals – such as Amnesty International – have argued that this narrow national definition fails to recognise the harm caused by hate speech based on other characteristics (Amnesty International, 2017). For instance, hateful comments directed at someone because of their gender, sexuality, religion or disability, may have a substantial and adverse impact on their mental health and wellbeing, and may contribute to wider patterns of discrimination and marginalisation. As a consequence, some – including, but not limited to the United Nations' Committee for the Elimination of Racial Discrimination – suggest that Section 18C should be strengthened (Koziol, 2017). There is a diversity of perspectives on various ways that this could be achieve, including:

- protect marginalised groups from discrimination and harassment;

- strengthen social cohesion through promoting tolerance and respect for diversity; and

- support dialogue and understanding between different groups.

However, some politicians, legal experts, and civil liberties advocates have controversially argued that legislative amendments must be enacted so as to weaken the application of Section 18C (Murphy, 2017). Some organisations and individuals – such as Senator Pauline Hanson, political commentator and former Senator Cory Bernardi and the Institute of Public Affairs – have argued that Section 18C restricts the freedom of speech and expression, and has an overall chilling effect on public debate (particularly around issues related to race and ethnicity) by limiting the ability of individuals to

express their opinions. Their main argument is, that in a representative and responsible democracy, individuals should be able to express their beliefs even if they offend or insult others, and that citizens should not be hesitant to express their views or engage in robust discussions on these topics for fear of being accused of racial discrimination.

Balancing the right to freedom of expression with the need to protect individuals and groups from the harms caused by hate speech is a complex policy and legislative challenge that requires careful consideration and consultation with a range of stakeholders. The Human Rights Commission acknowledged in its submission that freedom of expression is a fundamental human right – although Australians have only an implied right in this regard – but also notes that this right is not absolute and must be balanced with other human rights (such as the right to live with dignity and free from discrimination, and the right to security of the person).

The argument around freedom of expression plays a disproportionate role in blocking new regulation around hate speech – and the weaponisation of the argument is used by media personalities and companies. The Bolt vs Eatock case discussed above, where the Federal Court found Bolt's comments on light-skinned Aboriginal people breached the Racial Discrimination Act, demonstrates that effective legislation enacted through the courts does not necessarily align with public debate. Gelber and McNamara summarised this issue better than most:

> **"** The [publisher] did not agree with the substantive ruling, but decided to pursue its ongoing grievances by its own means, not in the courtroom (where the grounds of appeals would have to make specific and accurate reference to alleged errors in Justice Bromberg's reasoning). Bolt and [the publisher] had much greater freedom to frame the terms of the debate outside the courtroom and on the pages and websites of the Herald Sun and The Australian. They framed the main issues in a way that inaccurately characterised the decision and its effects, and positioned Bolt as the victim whose rights were violated instead of as the violator of the rights of others.
>
> (Gelber & McNamara 2013)

Despite there being broad and comprehensive support for protecting freedom of expression, there is a need for greater awareness initiatives to foster a better understanding of its limitations – particularly in acknowledging that, without better balance with other fundamental rights, we will continue to see a significant

detrimental impact on social cohesion, not to mention the actual harm caused to targeted communities and individuals.

2.2 Inconsistencies Between News and Social Media Regulation

The lack of consistency in how hate speech is regulated across different media platforms has been a source of concern for many experts and commentators. Social media platforms are not subject to the same level of regulation as traditional news media outlets. This is partly because social media platforms are not considered as publishers of content, but rather, as platforms that allow users to publish their own content. However, this perspective is increasingly weakening as more is understood about the algorithmic amplification of content by the platforms, which has been repeatedly shown to prioritise sensational, outrageous and conspiratorial content in order to fuel user engagement, regardless of the harm caused.

The Voller decision discussed in Section 2.1.3 highlighted some critics' concerns that placing responsibility with news media companies to moderate the comments section on their social pages, limits accountability of social media platforms. This is despite users being directed to those stories by the platforms' algorithms, which are also responsible – and arguably accountable – for elevating 'top comments' based on engagement, without regard for the harmful content they contain.

This is despite users being directed to those stories by the platforms' algorithms, which are also responsible – and arguably accountable – for elevating 'top comments' based on engagement, without regard for the harmful content they contain.

Social media platforms use their own self-set community standards and policies and are not subject to the same level of regulations and standards which govern traditional media outlets – including those set by ACMA – nor the same level of scrutiny and enforcement.

This makes it difficult to enforce standards of content and conduct across multiple forms of media, and it has led to concerns about the proliferation and dissemination of hate speech and other harmful content. Experts, including Dr Emma Jane from the University of New South Wales, argue that social media platforms and other online spaces require a specific set of regulations that take into account the unique features of online spaces,

such as the speed and scale at which information can spread, as well as the potential for anonymous communication.

A common factor in the regulation of news media and social media companies is the focus on their content rather than their systems, business models and algorithms that play a significant role in amplifying and inflaming dehumanising hate speech content. This focus produces a regulatory 'whack-a-mole' effect, in relation to some of the issues raised in this report:

> **Deficit discourse:** The current regulatory focus on explicit content means that it fails to detect dehumanising deficit discourse. This is despite deficit discourse helping to foster explicitly hateful comments and shaping long-term perspectives – exacerbated by the spread of comments based on engagement.

> **Mutualism:** There has been much written on the need to regulate social media's algorithms to reduce their amplification of sensationalist and emotional content. However, the incentives this system creates for news media companies to elicit online engagement place journalists in a difficult position, as it is clear that some factual headlines and content will be outperformed by sensational alternatives. With weak regulation and oversight, generating more engagement and outperforming competitors' content is seen as an obvious commercial imperative.

2.3 Enforcement Limitations

One of the major challenges in curbing dehumanising hate speech in Australia is the difficulty of enforcing regulations effectively with transparency and accountability. While there is a natural boundary between the nature of social media and news media and the required regulatory frameworks, the entwined relationships of their content and business models creates immediate regulatory gaps.

Enforcement is particularly difficult in regard to companies that are based overseas, enabling them to ignore notices from the eSafety Commissioner or other Australian regulators. This also means it is hard to mitigate the harms caused by the mutualistic relationship that exists between them and other hateful actors. Facebook and TikTok, two of the most widely-used social media platforms in Australia, are headquartered in the United States. They have their own community guidelines and policies for regulating content, but these are not subject to the same level of scrutiny and enforcement as traditional media regulations. With almost all major news outlets maintaining accounts on these platforms there is a gap that leaves accountability for

their content in a regulatory grey zone.

Even when industry codes for news and social media have been developed in collaboration with, or under the approval of, the regulator, these codes are often weak, voluntary or lack the appropriate consultation and input from the community. This is further evident in the use of terms from commercial broadcasters' of 'community standards' or social media platforms' development of 'community guidelines' which lack accountability and enforcement. The discussion here on the near-complete ineffectiveness of the Australian Press Council should nullify the defence from the numerous news media companies that there is sufficient regulation in place, and make clear that this focus on self-regulatory mechanisms will continue to produce weak and ineffective frameworks that are woefully inadequate at addressing the issues laid out in this report.

2.4 Identification, Monitoring and Removal of Hate Speech Content

Another challenge is identifying and monitoring hate speech in real-time. Vast amounts of harmful content is created and shared every day. During the COVID-19 pandemic, there was a significant increase in online hate speech towards Asian-Australians, with many individuals using social media to express racist sentiments towards the Asian community and blame them for the spread of the virus (Gover et al. 2020). However, the speed of this allowed harmful content to spread and cause harm before action could be taken.

Another example can be found in the challenge of monitoring hate speech during the Christchurch Terrorist Attack in 2019. The shooter live-streamed the attack on Facebook, which then spread to other social media platforms. This led to a significant amount of hate speech and online abuse targeted towards the Muslim community in Australia and around the world (Convery, 2022). The sheer volume of shares and reposts made it difficult for social media companies and law enforcement agencies to control the spread of harmful content in real time (Mahtani, 2019).

Social media companies use algorithms and other automated tools to identify and flag instances of hate speech on their platforms (Walsh 2021). However, there are several technical challenges associated with this. For instance, hate speech can take many different forms, as demonstrated in this report, and can be highly contextual, making it difficult for automated tools to accurately identify all instances of hate speech. An algorithm may flag a post containing a racist slur, but may not recognise a post that contains a more subtle form of hate speech, such as the use of coded language.

This is particularly concerning as automated tools are not yet equipped to account for the different cultural contexts in which hate speech can occur. These tools may need to be adjusted to account for the fact that a phrase considered hate speech in one country may not be recognised as such in another. This failure to account for different cultural and linguistic contexts was highlighted during the 2022 Australian federal election (Reset.tech Australia, 2022).

In addition, a post may contain hate speech directed at a particular race or religion, but if that group is not explicitly mentioned, an algorithm may fail to recognise the post as hate speech.

These issues are further exacerbated by the inability of researchers and civil society to monitor social media platforms. The monitoring tools used in this report are unique but are still limited, due to the social media platforms intentional limitation of external analysis of harmful content on their platforms.

3. Opportunities Beyond Regulation

Opportunities Beyond Regulation

While the effective regulation of dehumanising hate speech, news media and social media needs to remain a key priority, this is unlikely to be effective by itself given the issue's cultural, systemic and technological contexts. This report has outlined policy and regulatory approaches, but there are also opportunities that can collectively be classified as 'strength-based approaches'.

Here, a strength-based approach is intended to mean building the awareness, resilience and empowerment of the stakeholders within our information ecosystem. This approach provides additional opposition to deficit narratives, and is built on the work of the Lowitja Institute, Australia's National Institute for Aboriginal and Torres Strait Islander Health Research (Fogarty, Bulloch, et al., 2018; Fogarty, Lovell, et al., 2018).

The approaches below are considered outside typical regulatory functions, however, some should be enabled and backed by legislation.

3.1 Moderation and Monitoring Tools

I. Comment Moderation Tools
Most digital platforms already deploy various techniques and tools for monitoring comments on their websites, but their design is usually unclear and not always readily accessible by other actors. Moderation tools for encouraging healthier dialogue on media websites can equip journalists and website hosts or moderators to do this more effectively. One example is Coral, designed primarily for journalists but which also caters for a broader user base. It allows commenters to identify journalists in discussion threads, mute worrying voices and choose how to sort comments, while enabling moderators to easily remove disruptive comments, feature the best comments, view commenter histories and provide feedback.

II. Tools for Identifying Dehumanising Discourse
Most tools which detect dehumanising hate speech rely on the use of explicit terms and speech, and are not designed to identify dehumanising speech that is embedded in discourse or that is cumulative. The Australian

Muslim Advocacy Network has developed a _Hate Actor Assessment Framework_ which aims to enable the identification of such actors using dehumanisation as the primary measure.

III. Tools that Incorporate Social Context into Detection of Hate Actors

Tools designed to analyse content beyond basic text inputs, by incorporating insights from a commenter's social network, and their use of deficit discourse language more broadly, can better equip researchers and moderators. Researchers at the Indian Institute of Information Technology have found the effectiveness of incorporating social context into detection methods significantly outperforms other ways of analysing content.

But training and support is lacking for those eager to align with community expectations.

3.2 Education and Training for Journalists

The ethical and professional responsibility of journalists is key in countering deficit discourse, but training and support is lacking for those eager to align with community expectations.

I. Tools and Resources to Bridge Divisive Narratives

Journalists can unintentionally create 'us-versus-them' narratives which can inflame division and fuel deficit discourse online. More resources for journalists to understand the impact of these narratives is needed, and some are already available like the work of More in Common which is currently focused on the UK, US, France, Germany and Poland.

II. Making Better Assessments of Hate and its Impact

Journalists, under pressure from deadlines and the demand to generate engaging content, need tools to rapidly assess quotes and language for hatefulness and potential unintended harm. One example is the 5-point hate speech test from UK based Ethical Journalism Network, but more are needed.

3.3 Research and Knowledge Coalitions

Effective approaches to this issue will require more collaboration across news media and civil society. This will enable better approaches to storytelling, to undermine and prevent deficit narratives. Such partnerships already exist but there is a need for better resourcing for these groups, which are typically based within academic institutions or as not-for-profit entities. Some key examples:

- [Centre for Digital Wellbeing](#) - The Centre researches the effects of social media on mental health, misinformation and social cohesion and also makes submissions and recommendations for better policy and regulation.

- [News and Media Research Centre](#) - Australia's nationally recognised research centre for the study of news media industries, audiences and public discourse, conducts research and advocates for a media system that builds trust, inclusivity and diversity.

- [Tackling Hate](#) - This initiative at Deakin University adopts a 'whole-of-society' approach to addressing online hate. It collaborates with civil society organisations, Government agencies, law enforcement and academics to develop training on online hate analysis skills.

- [Online Hate Prevention Institute](#) – This aims to reduce the risk of people committing suicide, self-harming, or becoming involved in substance abuse, physical abuse and emotional abuse resulting from online hate. It has raised awareness of hate speech and its work has led to the removal of hate speech from several online platforms. Its [Fight Against Hate](#) software enables users to report instances of hate speech to a database independent of social media platforms which is used to inform researchers, law enforcement agencies and other stakeholders.

- [Chequeado](#) - An independent and non-partisan NGO that works to develop new tools and civic data sets in order to improve the quality of public debate and strengthen the democratic system through fact-checking, training and providing resources for journalists and the public.

4. Recommendations

4. Recommendations

Given that current regulation is ill equipped to tackle the patterns of dehumanising hate speech outlined in this report, and that it has gaps in its approach, we make a number of recommendations intended for governments, as well as for advocates of healthier online environments. Our recommendations include:

1. further research;

2. strengthen hate speech laws and balance freedom of expression;

3. address patchwork of news media regulation;

4. broaden diversity of news media ownership;

5. impose greater accountability on social media companies;

6. greater investment in media literacy programs; and

7. support and expand educational programs for news media practitioners

1. Further Research

There is a need to strengthen the evidence base for policy and program design, in order to deepen our understanding of the nature of dehumanising hate speech issues in Australia and other communities and countries. In addition to exploratory case study based research and quantitative approaches, more community-based reporting is needed, using interviews and participatory methods. These studies should focus on:

- the patterns of online hate and their impacts on specific communities;

- specific tactics or spaces that host online hate activity; or,

- the roles of certain kinds of actors within our information ecosystem.

This should include research to further uncover how different kinds of news media practitioners, from reporters to editors and social media production teams, view the online activity that surrounds their work, as well as focused studies with policy makers and practitioners in areas such as health and education, to assess how they experience online hate. We also recommend targeted, comparative analysis to assess how different styles and forms of

news media publishing correspond with the presence of hate speech. This could be extremely useful in training news media practitioners, in media literacy program design, and for policy makers.

2. Strengthen Hate Speech Laws & Balance Freedom of Expression

Online speech is only partially covered by the combination of Federal and State legislation, media and online platform regulators, and common law. Collectively they provide a range of definitions for types of speech people may seek recourse for, as well varying accounts for who is protected under those regulations. Effective hate speech legislation is the backbone of a holistic approach to countering dehumanising hate speech. The limitations of existing laws are demonstrated clearly in the case studies presented in this report. Greater certainty and clarity in this legislation will help inform other forms of regulation while also providing stakeholders with clearer guidelines and confidence in designing their programs and platforms appropriately.

A key component of this recommendation rests in bringing clarity to the balance between freedom of expression and protecting individuals from hate speech. Critical to this is acknowledging that dehumanising hate speech is socially contextual and that not all speech is equal. Regulations – both through hate speech laws and industry codes of practice – need to recognise this by taking into consideration the power dynamics and context in which it is produced. Gelber references this and the work of other experts in stating:

> the capacity to harm can derive from the social context within which an utterance is made. [McGowan] argues that because 'oppression is a social arrangement', some people are systematically disadvantaged in forms of social oppression (such as people of colour, women and LGBTQI folk), whereas others (such as those with bushy eyebrows) are not.
>
> (Gelber, 2019)

These needs outlined below build upon work from the Australian Muslim Advocacy Network (2023):

- There needs to be updated national definitions for dehumanising speech and discourse enabling hate and vilification. This may be through the expansion of the scope of Section 18C of the Racial Discrimination Act to provide:

a. new consistent definitions of hate speech including repeated deficit discourse narratives; and

b. broadening the definition of groups that are covered by the Act, to cover the identities and characteristics that actually inform the targeting of hate speech and vilification; that is, race, sex, age, disability, medical/health diagnoses (like, but not limited to HIV/AIDS), sexuality, gender identity, intersex, religion.

- Hate speech and disinformation needs to be included within the online safety framework, including civil penalties for dehumanising speech or discourse.

- Explicit rules should be given about the balance between freedom of expression and other fundamental rights, with an emphasis on evidence-based reasoning to consider the social context and to address unqualified freedom of speech claims.

- It should be clarified that Australia's discrimination and vilification laws apply to social media companies based overseas.

3. Address Patchwork of News Media Regulation

The patchwork of regulation that applies to news media is woefully inadequate in addressing the issue and a refreshed approach should establish the practice of codes and standards that avoid self-regulation.

There is a need to expand and update the powers of the media regulator to deal effectively with the complexity of modern online publishing. This will require greater resourcing to build the investigative powers of teams along with the expertise to deal with the wide range of online platforms including non-English language publishers, such as the Australian Financial News (Case Study 2).

Industry standards, including anti-dehumanisation ones, should be drafted by regulators with community and expert consultation, to inform more effective investigations and sanctions delivered by ACMA and eSafety.

4. Broaden Diversity of News Media Ownership

Governments, institutions and philanthropists need to dramatically increase their support for a broader diversity of news media publishers. This report clearly showed that News Corp, Australia's most dominant mainstream news

media company in terms of circulation, has been disproportionately represented in hate speech findings. This mirrors a pattern found across previous research that often singles out News Corp for its platforming of misinformation and negative, racialised portrayals of certain communities. The sheer size and dominance of the company may, in part, account for the disproportionate number of instances uncovered across research initiatives, however, it is well established that news media monopolisation leads to a lack of accountability and an outsized level of influence in dictating public narratives, and in shaping political debate and industry practice. Greater diversity of news media ownership will require:

Governments, institutions and philanthropists need to dramatically increase their support for a broader diversity of news media publishers.

- stronger media ownership diversity laws and antitrust legislation from governments; and

- government incentives and philanthropic investment in locally accountable and community-based news outlets to provide healthier competition.

5. Impose Greater Accountability on Social Media Companies

There is growing pressure on social media companies to be more proactive in preventing hate speech on their platforms. This could involve imposing greater accountability through, for example, imposing fines or other penalties if they fail to respond adequately. In a relative vacuum of regulation, there are numerous priorities that should inform this approach:

1. **Transparency and algorithmic audits** that provide access to regulators and researchers ('third-party audits') to analyse the outputs of algorithmic recommendation systems. Algorithmic audits have been set out in the EU's Digital Services Act, although Australian legislation should ensure this is not limited to audits conducted by social media platforms alone.

2. **Accountability for amplifying harmful content and comments** - While news media companies may have an implied responsibility for moderating comments on their social media pages, there should be some level of accountability for the platforms themselves in both hosting and amplifying harmful content and comments.

3. **Establishment of a positive 'statutory of duty care'** from social media platforms to users of their service. Perrin and Woods outlined this for the Carnegie Trust in a very straightforward way – that the duty of care 'would be in relation to the technology design and the operation of the platform by its owners, including the harm reducing tools available to protect its users and the enforcement by the platform of its terms and conditions. Parliament and the regulator would set out a taxonomy of harms that the duty of care was intended to reduce or prevent' (Perrin & Woods, 2018).

6. Greater Investment in Media Literacy Programs

While this report outlines the need to shift responsibility for detecting and countering hate speech from individuals, there will always be some burden placed on audiences in reporting and mitigating the issue. The Australian Media Literacy Alliance (AMLA) is a collection of key public institutions and networked organisations that defines media literacy as 'essential for full participation in society', and sets the needs across five key initiatives:

1. advocacy for high quality media literacy education;

2. world leading evidence-based research to underpin media literacy education;

3. national leadership through a network of media literacy champions;

4. national framework for teaching and measuring media literacy; and

5. production and circulation of engaging media literacy resources for all Australians.

This requires investment by governments and industry to facilitate this at the scale needed. Importantly, AMLA also states that media literacy is complemented by digital literacy and eSafety knowledge which provides important additions to the above initiatives. Those additions should include education programs and campaigns to build awareness of the relevant legislation and industry codes of practice, as well as digital literacy programs that focus on understanding the design of platforms and their algorithms. This helps to ensure communities and individuals are better equipped to understand, identify and report dehumanising hate speech – as well as knowing their rights in relation to this.

7. Support and Expand Educational Programs for Journalists and News Media Practitioners

Support and expand educational programs for journalists and other news media professionals on the impact that their work has when it reproduces or provokes the amplification of deficit discourse (see Key Definitions), dehumanisation and hate speech.

Appendices

Appendix A: AMAN Definition of Dehumanising Material

1. Dehumanising material is the material produced or published, which an ordinary person would conclude, portrays the class of persons identified on the basis of a protected characteristic ("class of persons") as not deserving to be treated equally to other humans because they lack qualities intrinsic to humans. Dehumanising material includes portraying the class of persons:

    a. to be, or have the appearance, qualities, or behaviour of

        i. an animal, insect, filth, form of disease or bacteria;

        ii. inanimate or mechanical objects; or

        iii. a supernatural alien or demon.

    b. are polluting, despoiling, or debilitating an ingroup or society as a whole;

    c. have a diminished capacity for human warmth and feeling or to make up their own mind, reason or form their own individual thoughts;

    d. homogeneously pose a powerful threat or menace to an in-group or society, posing overtly or deceptively;

    e. are to be held responsible for and deserving of collective punishment for the specific crimes, or alleged crimes of some of their "members";

    f.   are inherently criminal, dangerous, violent or evil by nature;

    g.   do not love or care for their children;

    h.   prey upon children, the aged, and the vulnerable;

    i.   was subject as a group to past tragedy or persecution that should now be trivialised, ridiculed, glorified or celebrated;

    j.   are inherently primitive, coarse, savage, intellectually inferior or incapable of achievement on a par with other humans;

    k.   must be categorised and denigrated according to skin colour or concepts of racial purity or blood quantum; or

    l.   must be excised or exiled from public space, neighbourhood or nation.

2.   Without limiting how the material in section (1) is presented, forms of presentation may include,

    a.   speech or words;

    b.   the curation or packaging of information;

    c.   images; and

    d.   insignia.

(Australian Muslim Advocacy Network, 2023)

Appendix B: Additional definitions of hate speech and related concepts

Further resources and definitions of hate speech, in addition to the definition of Hate Speech above under Key Definitions, include:

- United Nations hate speech definition

- Australian Human Rights Commission racial hatred defined

- Twitter hateful content definition

- Meta hate speech definition

- TikTok hateful behaviour definition

- Council of Europe definition and policy recommendations

- Australia Media, Entertainment and Arts Alliance (MEAA) Guidelines on Reporting Hate Speech and Extremism

Appendix C: List of Australian News Media Outlets

For those unfamiliar with the Australian news media landscape, a non-exhaustive list of popular Australian Media outlets in includes:

Television networks: Public televisions: ABC, SBS, NITV; Commercial televisions: Seven Network, Nine Network, Network 10, Sky News Australia, WIN Television.

Newspapers: The Australian Financial Review (national), The Australian (national), Herald Sun, The Canberra Times, Daily Telegraph, Sydney Morning Herald, Northern Territory News, The Courier-Mail, The Adelaide Advertiser, The Mercury, The Age, The West Australian, the Mercury.

News websites: News.com.au, ABC News, nine.com.au, 7news.com, Guardian Australia, Sydney Morning Herald, The Age, Herald Sun, Daily Telegraph, The Australian, Courier Mail, Brisbane Times, Adelaide Advertiser, Perth Now, WA Today, Canberra Times, Australian Financial Review, Daily Mail Australia, Junkee, The Saturday Paper, Sky News, The West, Sydney Today and Crikey.

News radio stations: ABC Radio National, Triple J (National), 2GB (Sydney), 3WA (Melbourne), 6PR (Perth) and 4BC (Brisbane).

References

Amnesty International.
(2017, January 18). Protecting Australians from Racial Discrimination. https://www.amnesty.org.au/racial-discrimination-section-18c-inquiry/

Alcorn, G. (2014, March 29). Locked in a war of words to define free speech. Sydney Morning Herald. https://www.smh.com.au/national/locked-in-a-war-of-words-to-define-free-speech-20140328-35oi1.html

Anti-Discrimination NSW. (2021, October 13). Vilification. Anti-Discrimination NSW; NSW Department of Communities and Justice. https://www.antidiscrimination.nsw.gov.au/anti-discrimination-nsw/discrimination/vilification.html

ATNS Project. (2020, December 14). Eatock v Bolt [2011] FCA 1103. Agreements, Treaties and Negotiated Settlements Project. https://database.atns.net.au/agreement.asp?EntityID=8430

Australasian Legal Information Institute. (n.d.) Racial Discrimination Act 1975. http://www8.austlii.edu.au/cgi-in/viewdb/au/legis/cth/consol_act/rda1975202/

Australasian Legal Information Institute. (1998, April 29). Royal Commission into Aboriginal Deaths in Custody—Reports Homepage. AustLII. http://www.austlii.edu.au/au/other/IndigLRes/rciadic/

Australian Communications and Media Authority. (2018, August 8). Investigation report no. BI-363. https://www.acma.gov.au/sites/default/files/2019-09/BI-363-Sunrise%20docx.docx

Australian Communications and Media Authority. (2020, May 19). Investigation report no. BI-529. https://www.acma.gov.au/publications/2020-05/report/bi-529-investigation-report

Australian Communications and Media Authority. (2023, March 3). Industry codes of practice. Acma.Gov.Au; Australian Communications and Media Authority. https://www.acma.gov.au/industry-codes-practice

Australian Human Rights Commission. (2016, September 8). Race hate and the RDA. Australian Human Rights Commission. https://humanrights.gov.au/our-work/race-discrimination/projects/race-hate-and-rda

Australian Muslim Advocacy Network. (2022, October 24). Policy Brief – Dehumanisation. Aman.Net.Au. https://www.aman.net.au/?page_id=1425

Australian Muslim Advocacy Network. (2023). Why Focus on Dehumanisation? Policy Brief, Dehumanisation. http://www.aman.net.au/?page_id=1425

Australian Press Council. (n.d.-a). Advisory Guidelines. Australian Press Council. Retrieved June 5, 2023, from https://presscouncil.org.au/standards/advisory-guidelines

Australian Press Council. (n.d.-b). "Asylum seekers", "illegal immigrants" and entry without a visa. Australian Press Council. Retrieved June 8, 2023, from https://presscouncil.org.au/document/asylum-seekers-illegal-immigrants-and-entry-without-a-visa

Burrowes, T. (2010, February 4). Media Watch host Holmes: unlike ACMA and the APC we can be brutal. Mumbrella. https://mumbrella.com.au/media-watch-returns-to-take-brutal-aim-at-offenders-in-the-media-17328

Byrne, E. (2021, September 8). High Court finds media outlets are responsible for Facebook comments in Dylan Voller defamation case. ABC News. https://www.abc.net.au/news/2021-09-08/high-court-rules-on-media-responsibility-over-facebook-comments/100442626

Clark, T. (2013, March 20). Media reform could save victims of misinformation … like my grandfather. The Conversation. https://theconversation.com/media-reform-could-save-victims-of-misinformation-like-my-grandfather-12909

Commercial Radio Australia. (2017). Commercial Radio Code of Practice. Commercial Radio Australia. https://www.commercialradio.com.au/CR/media/CommercialRadio/Commercial-Radio-Code-of-Practice.pdf

Convery, S. (2022, March 14). Muslims in Australia experienced surge of hate after Christchurch massacre, report reveals. The Guardian. https://www.theguardian.com/australia-news/2022/mar/15/muslims-in-australia-experienced-surge-of-hate-after-christchurch-massacre-report-reveals

Davis, L. P. M., & Museus, S. D. (2019). What Is Deficit Thinking? An Analysis of Conceptualizations of Deficit Thinking and Implications for Scholarly Research. NCID Currents, 1(1), 117–130. http://dx.doi.org/10.3998/currents.17387731.0001.110

Del Vicario, M., Zollo, F., Caldarelli, G., Scala, A., & Quattrociocchi, W. (2017). Mapping social dynamics on Facebook: The Brexit debate. Social Networks, 50, 6–16. https://doi.org/10.1016/j.socnet.2017.02.002

Dye, J. (2019, December 12). Seven settles with Aboriginal group over ‹Stolen Generation› Sunrise panel. Sydney Morning Herald https://www.smh.com.au/culture/tv-and-radio/seven-settles-with-aboriginal-group-over-stolen-generation-sunrise-panel-20191212-p53jd7.html

Fforde, C., Bamblett, L., Lovett, R., Gorringe, S., & Fogarty, B. (2013). Discourse, Deficit and Identity: Aboriginality, the Race Paradigm and the Language of Representation in Contemporary Australia. Media International Australia, 149, 162–173.

Fogarty, W., Bulloch, H., McDonnell, S., & Davis, M. (2018). Deficit Discourse and Indigenous Health: How narrative framings of Aboriginal and Torres Strait Islander people are reproduced in policy. Lowitja Institute.

Fogarty, W., Lovell, M., Langenberg, J., & Heron, M.-J. (2018). Deficit Discourse and Strengths-based Approaches: Changing the narrative of Aboriginal and Torres Strait Islander health and wellbeing. Lowitja Institute.

References

Free TV Australia. (2015). Commercial Television Industry Code of Practice 2018. Free TV Australia. https://www.freetv.com.au/wp-content/uploads/2019/07/Free_TV_Commercial_Television_Industry_Code_of_Practice_2018.pdf

Gagliardone, I., Patel, A., & Pohjonen, M. (2014). Mapping and Analysing Hate Speech Online (SSRN Scholarly Paper No. 2601792). https://doi.org/10.2139/ssrn.2601792

Gelber (2019). Differentiating hate speech: a systemic discrimination approach, Critical Review of International Social and Political Philosophy, 24 (4):1-22

Gelber (2021, July 14). A better way to regulate online hate speech: require social media companies to bear a duty of care to users. The Conversation https://theconversation.com/a-better-way-to-regulate-online-hate-speech-require-social-media-companies-to-bear-a-duty-of-care-to-users-163808

Gelber, K. & McNamara, L. (2013, December 11). Freedom of speech and racial vilification in Australia: 'The Bolt case' in public discourse. Australian Journal of Political Science, 470–484. https://www.tandfonline.com/doi/abs/10.1080/10361146.2013.842540?tab=permissions&scroll=topGherardi, S. (2019). How to Conduct a Practice-Based Study: Problems and Methods (2nd ed.).

Gover, A.R., Harper, S.B. & Langton, L. (2020). Anti-Asian Hate Crime During the COVID-19 Pandemic: Exploring the Reproduction of Inequality. American Journal of Criminal Justice, 45, 647–667

High Court of Australia. (2021). Fairfax Media Publications Pty Ltd V Dylan Voller Nationwide News Pty Limited V Dylan Voller Australian News Channel Pty Ltd V Dylan Voller [2021] HCA 27. High Court of Australia. https://cdn.hcourt.gov.au/assets/publications/judgment-summaries/2021/hca-27-2021-09-08.pdf

Ho, C. (2017, December 14). Racist reporting still rife in Australian media. The Conversation. http://theconversation.com/racist-reporting-still-rife-in-australian-media-88957

Jiang et al. (2020). Characterizing Community Guidelines on Social Media Platforms, CSCW ‹20 Companion: Conference Companion Publication of the 2020 on Computer Supported Cooperative Work and Social Computing. 287–291 https://dl.acm.org/doi/10.1145/3406865.3418312

Kerins, S. (2012). Caring for Country to Working on Country, People on Country: Vital Landscapes, Indigenous Futures; 26. The Federation Press.

Koziol, M. (2017, December 9). UN warns racism on rise in Australia, calls for section 18C to be strengthened. Sydney Morning Herald. https://www.smh.com.au/politics/federal/un-warns-racism-on-rise-in-australia-calls-for-section-18c-to-be-strengthened-20171209-h01qi0.html

Lomas, N. (2021, November 4) UK warns Facebook to focus on safety as minister eyes faster criminal sanctions for tech CEOs. TechCrunch https://techcrunch.com/2021/11/04/uk-warns-facebook-over-safety-criminal-liability-legislation/

Mahtani, S. (2019, March 17). Facebook removed 1.5 million videos of the Christchurch attacks within 24 hours - and there were still many more. The Washington Post.

https://www.washingtonpost.com/world/
facebook-removed-15-million-videos-of-the-
christchurch-attacks-within-24-hours--and-
there-were-still-many-more/2019/03/17/
fe3124b2-4898-11e9-b871-978e5c757325_
story.html

Media, Entertainment & Arts Alliance. (2020,
March). MEAA Guidelines on Reporting Hate
Speech and Extremism. MEAA. https://www.
meaa.org/meaa-media/code-of-ethics/hate-
speech-guidelines/

Media, Entertainment & Arts Alliance. (2021, April
21). Journalists' union gives notice to quit
Australian Press Council. MEAA
https://www.meaa.org/mediaroom/journalists-
union-gives-notice-to-quit-australian-press-
council/

Meta. (2023). Community Standards, Hate Speech.
https://transparency.fb.com/en-gb/policies/
community-standards/hate-speech

Murphy, K. (2017, March 30). Senate blocks
government›s changes to section 18C of
Racial Discrimination Act. The Guardian.
https://www.theguardian.com/australia-
news/2017/mar/30/senate-blocks-
governments-changes-to-section-18c-of-
racial-discrimination-act

National Inquiry into Racist Violence in Australia.
(1991). Racist violence: Report of the National
Inquiry into Racist Violence in Australia. Human
Rights and Equal Opportunity Commission.

Perrin, W. & Woods, L. (2018, March 22). Harm
Reduction In Social Media – A Proposal.
Collective wellbeing Carnegie UK. https://
carnegieuktrust.org.uk/blog-posts/harm-
reduction-in-social-media-a-proposal/

Reset.tech Australia. (2022, May 17).Response to
Meta: Safeguarding the integrity of our Federal
election. https://au.reset.tech/news/response-
to-meta-safeguarding-the-integrity-of-our-
federal-election-1/

RMIT ABC Fact Check. (2021, April 13). Just
how large is Rupert Murdoch's reach within
Australian media? ABC News. https://www.
abc.net.au/news/2021-04-14/fact-file-rupert-
murdoch-media-reach-in-australia/100056660

Rudd, K. (2021). Media Diversity in Australia Report.
Senate Environment and Communications
References Committee. https://www.aph.gov.
au/Parliamentary_Business/Committees/
Senate/Environment_and_Communications/
Mediadiversity/Report

United Nations. (2023). What is hate speech? United
Nations; United Nations. https://www.un.org/
en/hate-speech/understanding-hate-speech/
what-is-hate-speech

Wall Street Journal. (2021). The Facebook Files,
a Wall Street Journal Investigation, https://
www.wsj.com/articles/the-facebook-
files-11631713039

Walsh, E. (2021, October 18). Facebook claims
it uses AI to identify and remove posts
containing hate speech and violence, but the
technology doesn›t really work, report says.
Insider. https://www.businessinsider.com/
facebook-ai-doesnt-work-to-remove-hate-
speech-and-violence-2021-10

Wilson, C. (2023, June 23). Parties unite to
deliver greater internet censorship powers
to government-appointed official. Crikey.
https://www.crikey.com.au/2021/06/23/
parties-unite-to-deliver-greater-internet-
censorship-powers-to-government-
appointed-official/