Statutory Review of the Online Safety Act 2021

Submission paper

RMIT University Digital Ethnography Research Centre (Professor Rob Cover, RMIT University and Dr Jennifer Beckett, The University of Melbourne)

Paper prepared 20 June 2024.

Digital Ethnography Research Centre

RMIT UNIVERSITY

Digital Ethnography Research Centre

RMIT UNIVERSITY

21 June 2024

We thank the Department of Infrastructure, Transport, Regional Development, Communications and the Arts (DITRDCA) for the opportunity to respond to the Statutory Review of the Online Safety Act 2021 Issues Paper (April 2024).

Issues of online safety and digital harms have shifted substantially in the past half-decade, with a substantial, verifiable increase in online harassment and abuse, hate speech, adversarial online behaviour, doxxing (the release of private information to a public audience), and pile-on behaviour. We have witnessed a very substantial increase in harms directed to adults by adults, with reduced faith among online communities that platform policies, terms of service, moderation and guidelines are genuinely up to the task of addressing current rates of abuse.

Experiencing hostility, incivility and hate speech in online settings has become commonplace for users of digital communication and social media over the past decade. For example, a 2021 Pew Research Centre study found that 41% of adults in the United States have experienced some form of online harassment—a tripling of the rate in five years—and 25% of adults experienced more severe forms of online abuse, such as threats, stalking, sexual harassment and image-based abuse (Vogels, 2021). An estimated 14% of Australian adults are subject to hate speech and substantially more to other forms of online hostility (eSafety Commissioner, 2020), while a survey of women conducted for Amnesty International found that one-third had experienced some form of online harassment (Amnesty International, 2018a). In China, a poll of more than 2,000 social media users found 40% had experienced online abuse, with 16% of victim-survivors experiencing suicidality as a result (Radio Free Asia, 2022). Although many Australians have demonstrated resilience in the face of an increasingly toxic digital communication environment, the impact of harms on individuals and the degradation of what is now our key communication framework has the unfortunate capacity to leave many individuals under-supported and to shape social interaction in ways broadly undesirable, adversarial and hostile.

The need for multi-sector intervention and prevention is therefore clear, and we argue that this includes active engagement with the issues at the levels of regulation, education (children and adults), perpetrator penalisation, interjurisdictional cooperation, enforcing platform policy, and increased wellbeing resources.

Our submission is based on expertise in people-oriented research and lived experience of digital and mobile participation and research on those employed in digital enterprises. The authors draw heavily on work completed for a number of major funded research projects, including:

- the Australian Research Council Discovery Project *Online hostility in Australian Digital Cultures* (DP230100870) conducted by Rob Cover (RMIT University), Catharine Lumby (The University of Sydney), Benedetta Brevini (The University of Sydney), Jennifer Beckett (The University of Melbourne) and Jay Thompson (RMIT University);
- a study conducted for DITRDCA, *Public Figures and Online Abuse* (2023) by Professors Rob Cover and Nicola Henry;

- a current study on scam communication hosted by the RMIT Digital Ethnography Research Centre; among others.

We have responded to select questions in the Review Issues Paper most relevant to our areas of expertise and to the findings from our various research projects.

In preparing this submission we recognise and pay respect to the Elders – past, present, and emerging – of the lands on which our researchers and colleagues live and work.

Rob Cover (RMIT University) and Jennifer Beckett (The University of Melbourne)

Contact for correspondence

Professor Rob Cover

Co-director, RMIT Digital Ethnography Research Centre (DERC)

[redacted]

How to cite this document:

Cover, R. and Beckett, J. (2024). *Response to Key Issues in Online Safety in Australia: Submission Paper to the Statutory Review of the Online Safety Act 2021*, prepared 20 June.

Contents

1. Question 1: Are the current objects of the Act to improve and promote online safety for Australians sufficient or should they be expanded?

The current objects of the Act are to (a) improve online safety for Australians, and (b) to promote online safety for Australians. These are both sensible and appropriate objectives for guiding a regulatory framework.

However, we recommend an expansion of the Objects to include:
(c) to support Australians affected by online harms, including vulnerable persons and groups,
(d) to generate interjurisdictional dialogue for cooperative approaches to prevention and intervention of digital harms.

These additional Objects are explained in our responses to the below questions, although we summarise here that our research indicates

- a lack of practical support for those most affected by digital harms and a distrust in platforms to provide support, remedy or intervention leaving a sense of 'aloneness' and self-management or self-seeking of support and wellbeing, hence our addition of (c);
- a need to overcome the natural blockage to practical remedy that the international digital environment provides; Australia has been a leader on online safety but a key recommendation from us is that there are opportunities here to legislate that leadership role in terms of building and encouraging interjurisdictional (rather than platform) regulatory cooperation to ensure the safety of Australians from harm perpetration by international actors, and the safety of others international from harm perpetrated by Australians.

2. Question 4: Should the Act strengthen an enforceable Basic Online Safety Expectations?

We have recently witnessed the disappointment in Australia attempting to enforce basic online safety expectations in regard to requiring X (formerly Twitter) to take down harmful content that originated in Australia, and this provides a key indicator of the need for greater powers to enforce the core expectations on service providers who are providing a service in Australia.

The focus of the actionable element of the expectations over the past three years has been on the provision of data and the encouragement of conforming to internationally-recognised standards of intervention. However, strengthening the powers to enforce expectations requires not that the Minister be empowered to determine those expectations *but* that they be more clearly legislated and subject to legislative review on a three-yearly cycle.

The Issues Paper raises key areas where the Basic Online Safety Expectations warrant possible amendment, and we agree with these, particularly the additional reforms required in terms of generative artificial intelligence (and its role in harm, including the harm of bias).

We note that the focus on the interests of the child is, of course, important, but that the expectations should be extended to adults.

Expectations on Australian users

While the basic expectations are applied to service providers, this review is an opportunity to consider basic expectations on users. That includes Australians who are perpetrators of digital harms.

For example, there is an opportunity for new legislation that empowers the Minister and/or Commissioner to develop basic guidelines on acceptable online behaviour for Australians as a *mechanism* to help reduce the perpetration of abuse, harassment, participation in volumetric pile-ons. There is also an opportunity to legislate for the requirement for basic user expectations to be part of Australian educational curriculum. This is about setting cultural expectations and behavioural norms amongst Australians and allowing for a situation where users are offered "social proof of what the underlying injunctive norms are" (Kessler et al., 2012, p. 141) amongst users.

Finally, there is an opportunity to apply penalties to perpetrators of harm as a means of shaping good behaviour, in much the way that penalties for smoking (tobacco products) in prohibited areas has been very successful in fostering substantial social change in this regard. We address some of recommendations on penalties under the next question.

3. Question 6: To what extent should online safety be managed through a service providers' terms of use?

One of the key issues that has been highlighted in some of our research that has made it difficult to report, moderate, police and encourage good online behaviour in regard to the increase in abuse, harassment and pile-ons, has been a lack of an agreed code of behaviour in platform terms of service.

There is profound distrust among victim-survivors of online abuse and harassment that platforms will act on their terms of service or user guidelines in a way that is fair and transparent; indeed there is a belief among Australian and international participants in one of our studies that suggests that platform terms are not considered meaningful or useful, and that reporting instances of online abuse or harassment in regard to the terms is counter-productive (Cover 2022).

More significant, however, is that with large numbers of active platforms used daily by average individual users, the multiple and conflicting definitions across different platform terms is confusing, and leads to both difficulty reporting online harms to a platform, as well as not helping to shape perpetrators through a more generalised code of conduct for digital communication.

Finally, we note that our research has pointed to the fact that platform terms of service are among some of the worst documents in regard to *defining* the key elements they are purposed to address. For example, both the nature of what constitutes harm, abuse or harassment is often unclear or so broad that it is meaningless, and what constitutes

victimisation is virtually absent (Cover et al. 2024a). Differential thresholds for different types of users (e.g., public figures) exempt some users therefore from the same protections supposedly offered to other users, without clear reason or intent (Cover et al. 2024a). These issues all lead not only to a lack of public trust in the terms of service of the major corporate platforms, but to our expert opinion that they ineffective as tools to manage and reduce online abuse and harassment.

Volumetric pile-ons

Volumetric pile-ons are a good example of the shortcomings of platform terms of service, of automated filtering processes, and of extant regulatory practices. Pile-ons comprise very large number of participants joining in in criticism or shaming of a person, often a public figure or a user who has made a faux-pas or expressed a political viewpoint in the context of highly polarised political debate. Pile-ons can also be orchestrated by politically-active groups, often to shame or harm a member of a minority community. Pile-ons, however, are difficult to police in the current regulatory environment because they are comprised of individual, singular pieces of content that—in themselves—are usually very mild or mildly shaming (Thompson and Cover 2022). This leaves them liable to be ignored in platform reporting frameworks because *individual* pieces of content (post; replies) fall short of any threshold for intervention. However, in the context of the 'massified' ***effect*** of thousands of users piling-on one user, we see serious impacts beyond the known effects of singular instances of more serious content such as hate speech (Cover 2023). There is evidence of suicide subsequent to pile-ons (Thompson and Cover 2022); although further research is needed there is enough evidence to make a reasonable judgment that even well-intentioned participation in pile-ons can substantially harm another user's wellbeing and mental health.

This example provides us with an indicator that platform terms of service are inadequate for managing online safety, even if platforms responded to breaches more effectively and/or were regulated to do so.

Rather, what it points to is the need for intervention and prevention of volumetric pile-ons (and other known harms, abuses and forms of digital harassment) beyond reliance on platform terms and beyond regulation of platforms.

Some of our ethnographic work with Australian and international victim-survivors of online abuse and harassment indicates a public desire for stronger regulation of platforms, but also for:

- ***Education*** of users (including particularly adult users, e.g., through advertising) of the harms that may be caused to other users to help encourage users to understand that what is said online from a distance can be more than merely 'insulting' or 'offensive;
- ***Penalisation*** of perpetrators—as opposed to the penalisation of platforms who refuse take-down notices—to discourage uncivil behaviour online, including thoughtless participation in pile-ons, i.e., to require users to "rethink" their online social behaviour (Cover 2022).
- A greater public awareness of government ***complaints mechanisms*** including particularly the eSafety Commissioner's platform which, according to some of our nascent ethnographic work, is *virtually unknown* among those adult users who are most vulnerable or most in need of it.

In summary, [1] ***yes*** there should be further regulation of platforms' terms of service, including particularly:

a. a requirement for clear and standardised definitions across platforms (to guide both perpetrators and reporting practices); and
b. a removal of differential thresholds for public figures (since emergent public figures, and those who are under-supported such as individual influencers or family members of celebrities, are often captured by the poor definition of "public figure" in platform terms)

However [2], ***no*** terms of service should not reasonably considered a sole or primary mechanism for managing one's safety.

Legislation can require some of the cognate actions we have recommended above, including education, penalisation, advertising of third-party complaints mechanisms, etc.

4. Question 7: Should regulatory obligations depend on a service providers' risk or reach?

[and]

Question 19 What more could be done to enforce action against service providers who do not comply, especially those based overseas?

Interjurisdictional complexities have been at the heart of many of the difficulties faced by victim-survivors of online harms, including both the legal difficulties of enforcing take-down measures or other remedies across jurisdictions, and the strong likelihood of harms being perpetrated across three jurisdictions (the platforms', the perpetrator's, and the victim-survivor's) (Vincent 2017).

Indeed, our ethnographic work has shown not only that we cannot rely on all users, including victim-survivors seeking remedy, to understand inter-jurisdictional legal practices, but that very large numbers find the possibility of having to think through multiple jurisdictions so complex they are unwilling or unlikely to seek remedy and thereby experience protracted harms.

The contemporary socio-jurisdictional arrangement that assigns regulatory powers to nation-states (or the European Union) is at the core of the unworkability of some regulation of digital platforms (Suzor 2019).

This is where we recommend that there be legislation which empowers and funds the Minister and/or the Commissioner to lead the development of inter-jurisdiction and international bodies that can better help protect, remedy and govern digital platforms. Importantly, such a body could develop international standards, removing some of the inter-jurisdictional issues globally. Thinking back to the recent X vs eSafety Commission case, global standards of this may have forced the removal of the content by X.

There are extant models inter-jurisdictional policing and remedy-seeking that could be applied to the issue of online safety with a new, cross-jurisdictional and government-sponsored organisation.

For example, the International Criminal Police Organization (*Interpol*) was founded in 1923 to facilitate worldwide police cooperation.

Article 2 of the Interpol Constitution (1923) provides the relevant elements of its purpose:

- To ensure and promote the widest possible mutual assistance between all criminal police authorities within the limits of the laws existing in the different countries and in the spirit of the Universal Declaration of Human Rights.
- To establish and develop all institutions likely to contribute effectively to the prevention and suppression of ordinary law crimes.

As a starting-point for genuinely addressing the inter-jurisdictional issues of online safety and digital harms, and noting Australia's longstanding leadership on online safety, there are good grounds for requiring the Minister and/or Commissioner develop, encourage and frame the formation of an inter-jurisdictional body that can support the management of online safety investigation and issues resolution across jurisdictions.

Further, such an organisation would provide additional benefits in three respects:

1. There is a problem of unclear definition of terms in regard to online safety and digital harms across jurisdictions. For example, EU, UK and Australian legislation provides different meanings to different terms. This effective 'loophole' can and does allow both platforms and perpetrators to evade prosecution, and results in a failure to provide a common global 'language' and 'understanding' of digital harms and online safety, despite the setting for these issues being a remarkably globalised digital environment. An inter-jurisdictional body could provide and guide definitions and understandings in much the way the United Nations has been pivotal in guiding the definitions and determinations of 'hate speech' (broadly).

2. It would give users (globally, but also in Australia) additional options for remedy-seeking and greater confidence that there are bodies available that can help when facing an online safety issue, even if they do not seek that support.

3. In the longer term, an international body that works with digital platforms has the potential to provide greater pressure on digital platforms and other service providers who do not comply with local legislation or Commissioner requests in the context of our existing regulatory framework.

Broadly, the legislative review needs to recognise the difficulties of interjurisdictional practice, pressure and resolution that have been precisely highlighted over the takedown notice difficulties in 2024.

5. Question 8: Are the thresholds that are set for each complaints scheme appropriate? and Question 9: Are the complaints schemes accessible, easy to understand and effective for complainants?

We are in agreement with the Issues Paper's statement that content removal schemes can and do make a significant different to targeted individuals as a means of limiting harm (Part 3, p.19). The four complaints schemes are sensible, well-targeted and easy to understanding — although our ethnographic research indicates there are very large numbers of Australians who are unaware of these schemes. We would, of course, recommend that legislative pressure be used to advertise these better and more effectively.

We note, however, that one of the key issues in remedy-seeking of digital harms is an over-focus on *content* rather than *behaviour*. Although many platforms have a framework for reporting problematic content (such as harmful images) separate from problematic behaviour (such as harassment), we note that these are complex, not always well-understood and not necessarily always separable. The focus on *content* tends to allow harassment to be under-policed and under-remedied. Harassment and other problematic online behaviour often crosses multiple platforms, making a platform reporting basis resolution difficult and burdening the victim-survivor with substantial labour across different platforms with different terms of service.

Both the Australian complaints schemes and the reporting frameworks in the major commercial platforms thereby tend to obscure the sometimes greater harms of harassment and other problematic behaviour by focusing on specific, and identifiable, 'posts' or repeated content, rather that supporting the management and safety of those who are subject to harmful problematic behaviour of other users where those perpetrators' posts would not meet the threshold of problematic content.

In this respect, we recommend the legislation build towards better definitions and stronger complaints schemes that allow a user to seek resolution for problems of a perpetrator's behaviour that falls short of problematic content thresholds. There may be grounds, therefore, for separating the complaints of behaviour and content through a revision of the extant schemes, allowing users/complainants greater clarity on what they are reporting and in what context.

Additionally, we recommend a universal threshold for the current complaints scheme rather than the current two-tiered system of child vs adult, that has led to lower rate of successful complaints for adults vs children. It has also led to a system whereby a child who was bullied two-days before their 18th birthday, would meet the bar, but that same person two days later would face a bigger hurdle in having their complain upheld. While we completely agree that children do need protecting, the current settings make an assumption that harms are somehow lessened due to age, which is simply untrue. Harm to adults is often severe and can be complicated by a range of different issues such as underlying mental health concerns, socio-economic status, and relationship and family breakdowns and previous history, including childhood history of having experienced online abuse. This latter point is particularly salient amongst Millennials, Gen Z and future generations, all of whom have experienced growing up online to varying degrees.

6. Question 10: Does more need to be done to make sure vulnerable Australians at the highest risk of abuse have access to corrective action through the Act?

[and]

Question 27: Should the Commissioner have powers to act against content targeting groups as well as individuals?

Yes. It is widely recognised that migrant, ethnic, racial, gender and sexual orientation minorities are more likely to be targeted by online abuse and harassment and to experience harms more greatly than their non-minority peers. This can include the targeting of groups through extremist and intolerant perspectives, even when they do not address a vulnerable individual—the harms are nevertheless apparent and widely felt.

The UN Special Rapporteur on Minority Issues (2022) has noted that there is a tendency in both platform practice and public discussion to fail to balance the need to protect minorities and freedom of expression, often favouring the latter even when clear harms to minorities is likely.

There are very good grounds for this review to recommend amendments to the Act that enable an enforcement on both platforms and perpetrators where harms to minorities and other vulnerable users is likely. We would recommend not only empowering the Minister to determine some vulnerable groups (emergent minority identities and other vulnerabilities that may not always be tied to identity or community), but also to enshrine in legislation some key minority groups who are the most targeted and most likely to be harmed.

Outside of concerns for minority groups, another vulnerable population are those who work with social media behind the scenes. More work needs to be done to address the workplace harms of their exposure to online incivility and toxicity. This is a particular concern for workers in the growing fields of social media management, online community management, journalism, and those who work in moderation and remediation in organisations such as the Office of the eSafety Commissioner and, more recently workers tasked with cleaning and regulating datasets for training AI. Research, as well as several high-profile legal cases, has shown that continued exposure to this material can lead to PTSD (Spence et al, 2024). Anecdotal evidence from content moderators has also shown that the work of moderation can also be a pathway to radicalisation (Gray, 2022). Ensuring workplace safety for this group of people who work to keep social media spaces safer overall will have a net-benefit in the long-run.

7. Question 14: Should the Act empower ‘bystanders’, or members of the general public who may not be directly affected by illegal or seriously harmful material, to report this material to the Commissioner?

There is some evidence in scholarship that bystanders are more likely to intervene directly when they witness hate speech or other abuses or harassment than the victim-survivor (Obermeier et al. 2021), and this gives credence to the suggestion that the Act should

indeed empower bystanders who are not directly affected to make reports to the Commissioner through the extant reporting mechanism.

This may be more important than simply utilising bystanders to aid in reporting. Bystanders themselves are known to experience the impact of uncivil online behaviour or hate speech by witnessing in two forms:

1. Being exposed to high rates of toxic behaviour and suffering harms to health and wellbeing, particularly when it involves racialised hate speech not directed to them but to a group (Wachs and Wright 2018);
2. Experiencing disinhibition from exposure to online abuse and harassment which normalises the practice among a wider group of bystanders (Keighley 2022).

While, again, substantially more research is needed on the role and experience of bystanders who are neither victim-survivors of abuse/harassment online nor perpetrators, a reasonable argument can be made that by encouraging bystanders to make reports to the Commissioner may serve as an important mechanism for communicating in a more widespread way that online abuse and harassment are harmful, problematic and damaging to the wider digital ecology.

8. Question 18: Are Australia's penalties adequate and if not, what forms should they take?

The Act should be amended to follow the practice in Ireland, the European Union and the United Kingdom for substantially higher penalties for platforms that are non-compliant with regulation or regulatory requests, and the percentage of global revenue is a sensible framework in order not to unduly penalise under-resourced small or non-corporate platforms who may not be in a position to act as quickly as the major corporate platforms.

The challenges to enforcing penalties on individuals outside the jurisdiction is a serious one, given cross-jurisdictional harms are among the most likely and most common for some targeted groups. *See s4 above for our recommendations on interjurisdictional remedies to the penalisation and policy framework.*.

More importantly, the act does not provide detail for the penalisation of individual perpetrators of specific harms such as online abuse, harassment, orchestration of pile-ons, etc., with the focus on individuals who are responsible for a take-down. This is where we would like to recommend a substantial shift in practice. While it is widely recognised that the Act and the role of the Commissioner is to utilise regulatory mechanisms upon platforms, it is also widely regarded that some of the major platforms are not doing enough to ensure the safety of users, and that regulatory pressures are not working (Suzor 2019; Flew 2021).

We recommend an urgent need for a penalty regime for Australians who are perpetrators of online safety issues, abuses, harassment or other problematic behaviour to serve three purposes:

1. To shape normative online behaviours among Australian users in much the way financial penalties for smoking in certain public places has been highly successful in producing social change (Wynne et al, 2018)
2. To actively discourage Australians from perpetrating abuse or harassment online targeting other users both within and outside Australia; and
3. To provide a legal and penalties framework that can underpin advertising (for adults) and education (for young people) on the significance of perpetrating online harms, and the risks to themselves of damaging the digital ecology through persistent perpetration.

This is not to suggest that a civil penalties schema for Australian perpetrators is a wholesale solution to online harms, not to suggest that such a schema can be quickly and easily built. It opens a number of important questions, including:

- should perpetrators be penalised through bans of use of internet services or financial penalties only?
- how do we manage inequitable penalisation in regard to those with fewer social-economic resources vis-à-vis their better resourced peers?
- in what ways can a perpetrator appeal a penalty without putting pressure on magistrates' courts? who else can consider an appeal?
- should the penalties be high enough to cover the costs of a penalties framework?
- should penalties be jointly across platforms and users, in much the same way the NSW Smoke-free Environment Regulation 2016 applies penalties both to the owners of a premise where prohibited smoking is occurring and to the smoker themselves?

The questions are complex, but there is good reason based on the effectiveness of civil penalties in shaping other kinds of behaviours that amended legislation could participate in the shaping of the Australian user population for better behaviour, less incivility and greater positive social engagement to the benefit of the entire community.

9. Question 24: Should there be a mechanism in place to provide researchers and eSafety with access to data? Are there other things they should be allowed access to?

Yes. Although there is a great deal of research activity on digital harms and online safety, and a broad literature and scholarship on the topic, there are continued gaps on what constitutes digital harms, and filling these scholarly gaps requires that the Commissioner, university researchers, and community advocacy groups have greater access to platform data.

For example, half a decade ago, 'doxxing' (the illicit release of a person's private information or identifiers to a wide online group without their permission) was unknown as a harmful digital practice. Today, it is widely recognised, due in particular to some good journalism on the topic during doxxing scandals over the past three years. However, there are research gaps that warrant data acquisition from platforms to enable key insights on the extent of harm, the impact of harm, the resolution possibilities and the effectiveness of platforms in remedying.

We recommend the legislation empower the Commissioner to require broad de-identified research data from platforms and to enable the sharing of this data with universities and recognised community and advocacy groups; and that the data not be limited so as not to foreclose on unseen future needs.

10. Question 32: Does Australia have the appropriate governance structures in place to administer Australia's online safety laws?

With the exception of the responses made above in regard to changes to penalisation of perpetrators and interjurisdictional management of complaints, we are very supportive of the eSafety Commissioner and the role the Commissioner plays. We feel the role could be expanded into greater promotion of education and social 'shaping' of good online behaviour to help prevent the incivility and hostility that has become a normative experience of participating online.

There are some areas where a known lack of clarity across the governance structure exists: for example, the role of the Australian Communications and Media Authority (ACMA) vis-à-vis the Commissioner's office is unclear to many people, particularly in the areas of abuses (such as scams) via mobile technologies that fall under the purview of the ACMA.

Although we do not have the data to know if there are likely cost-savings, there may be benefits in considering the relationship of several extant bodies at the time of review of this legislation, and we would strongly recommend that be a consideration.

References

Amnesty International (2018a) Australia: poll reveals alarming impact of online abuse against women. *Amnesty International*, 7 February. Available at: https://www.amnesty.org.au/australia-poll-reveals-alarming-impact-online-abuse-women/ (accessed 12 March 2024).

Cover, Rob (2022). Digital hostility: Contemporary crisis, disrupted belonging and self-care practices. *Media International Australia* 184(1): 79-91.

Cover, Rob (2023). 'Digital Hostility, Subjectivity and Ethics: Theorising the Disruption of Identity in Instances of Mass Online Abuse and Hate Speech.' *Convergence: The International Journal of Research into New Media Technologies* 29(2): 308-321. DOI: 10.1177/13548565221122908.

Cover, Rob; Henry, Nicola; Gleave, Joscelyn; Greenfield, Sharon; Grechyn, Viktor and Huynh, Thuc Bao; (2024). 'Protecting public figures online: How do platforms and regulators define public figures?' *Media International Australia*. Online first. DOI: 10.1177/1329878X231225745

eSafety Commissioner (2020) *Online Hate Speech*. Available at: www.esafety.gov.au/sites/default/files/2020-01/Hate%20speech-Report.pdf (accessed 19 December 2023).

Flew, Terry (2021). *Regulating Platforms*. Cambridge: Polity.

Gray, Chris (2022). *The Moderator: Inside Facebook's Dirty Work in Ireland*. Dublin: Gill Books

Keighley, Rachel (2022). 'Hate Hurts: Exploring the Impact of Online Hate on LGBTQ+ Young People.' *Women & Criminal Justice*, 32(1-2): 29-48.

Kessler, S.E., Kraut, R.E., Resnick, P., and Kittur, A. (2012) Regulating behaviour in online communities, in: *Building Successful Online Communities : Evidence-Based Social Design*. Cambridge: MIT Press (pp. 125-178)

Obermaier, Magdalena; Schmuch, Desirée and Sasleem, Muniba (2021). 'I'll be there for you? Effects of Islamophobic online hate speech and counter speech on Muslim in-group bystanders' intention to intervene.' *New Media & Society*. Epub ahead of print: DOI: 10.1177/1461448211017527.

Radio Free Asia (2022) *The Chinese Internet's Hidden Victims: Uncovering and healing the scars of online abuse.* https://www.wainao.me/wainao-reads/uncovering-and-healing-the-scars-of-online-abuse-04132022 (accessed 9 March 2024).

Spence, R., Bifulco, A., Bradbury, P., Martellozzo, E., & DeMarco, J. (2024). Content Moderator Mental Health, Secondary Trauma, and Well-being: A Cross-Sectional Study. *Cyberpsychology, Behavior, and Social Networking*, *27*(2), 149-155.

Suzor, Nicolas P. (2019). *Lawless: The Secret Rules that Govern our Digital Lives*. Cambridge: Cambridge University Press.

Thompson, Jay Daniel and Cover, Rob (2022). 'Digital Hostility, Internet Pile-ons, and Shaming: A Case Study.' *Convergence: The International Journal of Research into New Media Technologies* 28(6): 1770-1782. DOI: 10.1177/13548565211030461.

UN Special Rapporteur on Minority Issues (2022). *Draft 'Effective Guidelines on Hate Speech, Social Media and Minorities.'* https://www.ohchr.org/sites/default/files/2022-06/Draft-Effective-Guidelines-Hate-Speech-SR-Minorities.pdf

Vincent, Nicole A. (2017). 'Victims of Cybercrime: Definitions and Challenges.' In Elena Martellozzo and Emma A. Jane (eds.), *Cybercrime and its Victims*. Abingdon: Routledge, pp. 27-42.

Vogels EA (2021) The state of online harassment. *Pew Research Center,* 13 January. Available at: https://www.pewresearch.org/internet/2021/01/13/the-state-of-online-harassment/ (accessed 12 September 2023).

Wachs, Sebastian and Wright, Michelle F. (2018). 'Associations between Bystanders and Perpetrators of Online Hate: The Moderating Role of Toxic Online Disinhibition.' *International Journal of Environmental Research and Public Health* 15(9): 2030.

Wynne, O., Guillaumier, A., Twyman, L., McCrabb, S., Denham, A. M., Paul, C., ... and Bonevski, B. (2018). Signs, fines and compliance officers: a systematic review of strategies for enforcing smoke-free policy. *International Journal of Environmental Research and Public Health*, *15*(7), 1386.

About the Online hostility in Australian Digital Cultures (ARC Discovery Project DP230100870)

This project aims to provide a comprehensive account of Australians' experiences of online hostility, abuse, trolling and extremist hate speech, which have increased over the past decade. The research expects to analyse the experiences of diverse Australian online users, moderators and stakeholders, to determine their practices, attitudes, and innovations, and their perceptions on how to address this social problem. Expected outcomes of this project include enhanced understanding of the support needs and remedies to online hostility among a diverse cross-section of Australians. This will provide significant benefits by providing roadmaps for improved intervention, support, regulation and education on digital communication in Australia.

Funding details

Australian Research Council Discovery Projects Scheme DP230100870 (2023-2026)

Chief investigators

Professor Rob Cover (RMIT University), Dr Jennifer Beckett (The University of Melbourne), Associate Professor Benedetta Brevine (The University of Sydney), Dr Jay Thompson (RMIT University) and Professor Catharine Lumby (The University of Sydney).

Other team members

Joscelyn Gleave (RMIT University), Joel Humphries (RMIT University), Rhyle Simcock (Queensland University of Technology).

Project publications

https://digitalhostility.org/publications/

About the RMIT Digital Ethnography Research Centre

The Digital Ethnography Research Centre (DERC) conducts research on the lived experience of digital and mobile technologies, cultures and adaptations, each of which is inextricable from the environments and relationships in which everyday life and work plays out.

DERC excels in both academic scholarship and in our applied work with external partners from industry and other sectors.

DERC approaches this world and how we experience it through innovative, reflexive and ethical ethnographic approaches, developed through anthropology, media and cultural studies, design, arts and documentary practice and games research, to provide people-centric meaningful data and analysis that informs policy, safety-by-design and social support in a fast-changing digital communication ecology.

We study everyday lived experience, the future of work, creative youth practices, tech policies, tech design, digital literacy and a range of other human-centred design propositions.

Hosted by RMIT in Melbourne, Australia, DERC offers world class training ground and expertise for both small and largescale research initiatives. We are the first and largest research centre in Australia focused on digital ethnography.

DERC convenes the *Digital Hostility and Disinformation Lab* which works to fosters cross-sector partnerships to address the social and ethical implications of contemporary forms of digital hostility, disinformation, dark participation, and platform misuse, all of which are toxifying the contemporary digital ecology.

About the authors

Professor Rob Cover

Rob is Professor of Digital Communication at RMIT University and Co-Director of the RMIT *Digital Ethnography Research Centre*. He researches youth wellbeing and identity in digital contexts. The author of 10 recent books including *Queer Youth Suicide, Culture and Identity: Unliveable Lives?* (2012); *Digital Identities: Creating and Communicating the Online Self* (2016); *Emergent Identities: New Sexualities, Gender and Relationships in a Digital Era* (2019), *Fake News in Digital Cultures* (with A Haw and JD Thompson, 2022) and *Identity and Digital Communication* (2023), he leads a number of ARC research projects on digital cultures and minorities, and works with government and industry to provide lived experience insights on digital cultures.

https://www.rmit.edu.au/contact/staff-contacts/academic-staff/c/cover-professor-rob

Dr Jennifer Beckett

Jennifer is senior lecturer in Media and Communications at The University of Melbourne. Her research primarily looks at online governance of social spaces on the internet, from large social media companies like Facebook to smaller online communities. Before joining the University of Melbourne, Jennifer worked as an online and social media producer for radio current affairs at the Australian Broadcasting Corporation.

https://findanexpert.unimelb.edu.au/profile/686177-jennifer-beckett