Combatting Online Islamophobia and Racism in Australia

The case for an eSafety duty of care

Umar Butler
In the spirit of reconciliation, we acknowledge the Traditional Custodians of Country throughout Australia and their connections to land, sea and community. We pay our respect to their Elders past and present and extend that respect to all Aboriginal and Torres Strait Islander peoples today.

Authored by
Umar Butler
On behalf of the
Islamic Council of Victoria

© Umar Butler and the Islamic Council of Victoria 2022

This report is licenced under a Creative Commons Attribution 4.0 International Licence, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author and the source, provide a link to the Creative Commons licence, and indicate if changes were made.

Cite this report as: Butler, Umar, Combatting Online Islamophobia and Racism in Australia: The Case for an eSafety Duty of Care (Report, Islamic Council of Victoria, 18 October 2022)
<https://doi.org/10.25916/4gtq-m054>
1 Introduction

As Islamophobia grows to ‘epidemic proportions’ around the world, social media platforms have correspondingly seen an explosion in hateful content. In the case of Twitter, for instance, findings from our recent report Islamophobia in the Digital Age: A Study of Anti-Muslim Tweets show that the number of Islamophobic posts made each year is now in the millions, with at least 2,129,861 anti-Muslim tweets being published in 2020 alone. More broadly, there are also indications that Islamophobia occurs with greater frequency on the internet than it does in real life. As we noted in our last report, this trend is ‘fuelling a vicious cycle whereby online expressions of hate incite offline attacks on Muslims, which in turn provoke more online hate’.

Despite the apparent severity of rampant digital Islamophobia, platforms have remained largely complacent, refusing to improve their demonstrably ineffectual systems for the review of hateful material. The Islamic Council of Victoria (‘ICV’) finds that this failure, coupled with the grave harms of online anti-Muslim hate, necessitates government intervention. While there are a number of competing approaches to the regulation of social media, our preference is to reform the systems that have enabled and, indeed, at times encouraged the widespread and unchecked dissemination of hate speech rather than attempt the practically impossible task of taking down hundreds

---

3 Ibid 9.
5 Butler (n 2) 4.
of millions of individual pieces of Islamophobic content. To implement this approach, we propose that Australia place a statutory duty on platforms to take reasonable care to protect users from harm (the ‘eSafety duty of care’), similar to the regulatory scheme set to be established by the UK’s Online Safety Bill 2021.

Cognisant of the risks posed to users’ freedom of expression by the regulation of social media, in addition to the many burdens that an eSafety duty of care would place on platforms, we nevertheless do not find such harms to outweigh those that would inevitably flow from a failure to curb the growth of online Islamophobia. Indeed, as the events of the Quebec City mosque shooting, Delhi riots and Christchurch attacks should teach us, the very real capacity for digital hate to transform into physical violence means that human lives are at stake every time a platform refuses to prevent the spread of Islamophobic material.

The corresponding structure of this report is as follows. Part 2 argues that platforms have failed to protect users from rampant online anti-Muslim hate. Part 3 makes the case for government intervention. Part 4 advocates for a regulatory approach aimed at reforming the systems that have enabled the unchecked growth of digital Islamophobia. Part 5 introduces the concept of an eSafety duty of care, exploring its reception in the UK and providing practical recommendations as to how it could be implemented in Australia. Part 6 concludes.

7 See generally Reset, Submission No OSH0024 to Commons Digital, Culture, Media and Sport Sub-Committee on Online Harms and Disinformation, UK Parliament (September 2021) [3]–[6] <https://committees.parliament.uk/writtenevidence/38548/pdf/>.
9 Online Safety Bill 2021 (UK) (‘UK OSB’).
2 Platforms Are Failing Users

Our report on anti-Muslim tweets provides some of the clearest evidence to date of the pervasiveness of Islamophobia on social media. In it, we find that there were at least 3,759,180 anti-Muslim posts shared on Twitter between 28 August 2019 and 27 August 2021, 85.17% of which remained online almost a year later, on 27 June 2022.

These discoveries were made by, first, designing a search query to capture as many Islamophobic tweets as possible (barring, due to practical limitations, non-English tweets, retweets and tweets containing photos, videos and GIFs), after which we collected 6,603,114 posts published two years from the day that our work commenced, with 10,000 being randomly sampled for manual classification as either Islamophobic or neutral. The definition of Islamophobia employed was taken from our *Islamophobia Position Statement*:

Islamophobia is ‘rooted in racism and is a type of racism that targets expressions of Muslimness or perceived Muslimness’. Very simply, Islamophobia is ‘anti-Muslim racism’. Islamophobia involves the ‘rejection of Islam, Muslim groups and Muslim individuals on the basis of prejudice and stereotypes. It may have emotional, cognitive, evaluative as well as action-oriented elements’.

Applying this definition to our sample, we used the resulting annotations to train a machine learning model to identify anti-Muslim hate with 86.68% accuracy. That model was then employed to automatically classify our dataset of 6,603,114 posts, resulting in the above-mentioned finding that there were at least 3,759,180 Islamophobic tweets made between 28 August 2019 and 27 August 2021. Finally, on 27 June 2022, we reran our original search query to uncover that a mere 14.83% of anti-Muslim posts had been removed (whether by users themselves or Twitter moderators).

13 Butler (n 2).
14 Ibid 7–8.
17 Butler (n 2) 7.
18 Ibid.
19 Ibid 8.
From these findings, several conclusions may be drawn about the prevalence and severity of digital anti-Muslim hate. First, in light of the fact that our search query was limited to keywords, phrases and combinations thereof with a higher-than-average likelihood to result in the collection of Islamophobic tweets, and also that we were unable to survey non-English tweets and tweets containing photos, videos and GIFs, we estimate the actual number of anti-Muslim posts made on Twitter each year to be in the tens if not hundreds of millions, rather than just millions. Furthermore, given that most other major platforms such as Reddit, YouTube and Facebook have been found to be just as poor as Twitter at responding to reports of hate speech, it may be inferred that they see relatively similar amounts of Islamophobic content.

The low ratio of posts that ended up being removed also evinces a dramatic failure on the part of Twitter to moderate anti-Muslim hate. This conclusion is supported by the Australian Muslim Advocacy Network, which recently lodged a complaint with the Queensland Human Rights Commission accusing Twitter of refusing to take down 419 vilificatory Islamophobic posts, including comments referring to the Qur’ān as a ‘terrorist handbook’ and Islam as the ‘most violent and sexually perverse cult’. Again, however, it is not just Twitter that is failing Muslim users. In a new study, the Center for Countering Digital Hate flagged 530 Islamophobic posts on Twitter, TikTok, Instagram, YouTube and Facebook to find that just 11.3% were acted upon, with YouTube scoring the worst by ignoring all reports and TikTok the best by removing 36% of hateful posts.

Considered together, the findings made by the ICV and others demonstrate both the rampancy of Islamophobia on social media and the corresponding complacency of platforms. As will be shown in the next section, when attention is also given to the grave harms of online anti-Muslim hate, it becomes apparent that government intervention is necessary.

22 Butler (n 2) 8.
24 Center for Countering Digital Hate (n 21) 10.
3 The Need for Government Intervention

Although the regulation of social media may very well risk limiting users’ freedom of expression as well as further imposing weighty legal and financial burdens on platforms, we do not view such detriments to outweigh the many severe effects already being had by pervasive digital Islamophobia on the mental and physical wellbeing of Muslims.

Research shows, for example, that exposure to online Islamophobia can, inter alia, diminish individuals’ confidence, erode their sense of identity and instil debilitating anxiety and dread.\(^{25}\) Indeed, in a study of 20 British victims of online anti-Muslim hate, Irene Zempi and Imran Awan found that some participants had become so wrought with fear that they began to question whether Islamophobes might attempt to harm them in real life.\(^{26}\) They quote one such victim as saying: ‘I am scared, I fear for my life because at the end of the day they [cyber perpetrators] might come and find me because my Twitter profile is public.’\(^{27}\) Another participant remarked: ‘I thought that they would turn up at my house, and hurt me and my family because they knew who I was and where I lived.’\(^{28}\)

Zempi and Awan’s findings on the devastating psychological and emotional impacts of anti-Muslim hate are further supported by Derya Iner’s latest Islamophobia in Australia report.\(^{29}\) Out of 109 cases of online Islamophobia, 59% led to feelings of anger, 49% sadness/worry, 18% disappointment, 8% fear and 2% humiliation.\(^{30}\) One individual who flagged a post pledging to kill 10 Muslims

\(^{26}\) Irene Zempi and Imran Awan, Islamophobia: Lived Experiences of Online and Offline Victimisation (Policy Press, 2016) 15, 58.
\(^{27}\) Ibid 58.
\(^{28}\) Ibid 59.
\(^{30}\) Ibid 128.
a day commented: ‘[The offender’s] public posts are extremely disturbing & I’m afraid he might cause harm. This post is not ok and action needs to be taken in addressing this hate speech & radicalism.’

Similarly, the reporter of a post arguing that it is rational to fear Muslims remarked: ‘[I] realised the extent of Islamophobia that still exists. Will be extra cautious in public.’

Of course, these are not the only effects digital Islamophobia can have on victims’ psyches. On 5 April 2020, a young Indian Muslim man committed suicide after being falsely accused of deliberately spreading the coronavirus (known as the ‘corona jihad’ conspiracy theory).

More broadly, a 2021 study shows that US Muslims are two times more likely to report a history of suicide attempt than other religious groups, including atheists and agnostics. While that statistic is likely the result of a multitude of causes, rampant online Islamophobia certainly does not help, and, indeed, is probably a contributing factor.

In addition to compromising Muslims’ mental wellbeing, digital Islamophobia is also ‘fuelling a vicious cycle whereby online expressions of hate incite offline attacks on Muslims, which in turn provoke more online hate’. The Delhi riots are a prime example. According to an internal Facebook report reviewed by The Wall Street Journal, the weeks and months preceding the February 2020 outburst of communal violence were marked by a 300% increase in inflammatory content, examples of which included ‘dehumanizing posts comparing Muslims to “pigs” and “dogs”’, ‘misinformation claiming the Quran calls for men to rape their female family members’ and assertions that ‘Muslim clerics spit on food to either “make it halal”, or spread Covid-19, as a larger war against Hindus’.

Then, during the riots themselves, social media was used to further escalate tensions through the viral dissemination of false rumours of a Muslim uprising and incendiary calls for Hindus to defend themselves.

We note, for instance, that in the first week of the riots (from 23 to 29 September),

31 Ibid 130.
32 Ibid.
34 Rania Awaad et al, ‘Suicide Attempts of Muslims Compared with Other Religious Groups in the US’ (2021) 78(9) JAMA Psychiatry 1041, 1041.
35 Butler (n 2) 4.
36 Purnell and Horwitz (n 11).
the number of Islamophobic tweets grew by a staggering 181%. The end result of this surge in anti-Muslim hate was the death of 53 and injury of over 200.

Further evidence of how digital Islamophobia can provoke physical violence and vice versa may be found in the Christchurch attacks. As we observed in our report on anti-Muslim tweets, ‘Not only was it online that [the perpetrator] was radicalised, but it was also there that he was able to livestream his massacre of 51 New Zealand Muslims, triggering an explosion in Islamophobic material and the corresponding brainwashing of countless new far-right extremists.’ It is perhaps also noteworthy that the attacker drew inspiration from the Quebec City mosque shooter, an individual who likewise happened to be radicalised by anti-Muslim and far-right content.

Both the grave psychological harms of online Islamophobia and its very real capacity to translate into physical attacks on Muslims have been known to platforms for close to a decade. As early as 2013, for example, the Centre for Fascist, Anti-Fascist and Post-Fascist Studies issued a report warning platforms that ‘online expressions of anti-Muslim sentiment have reached significant proportions’ and could be a ‘precursor to more physically threatening offline incidents’. Nine years later and the pleas of researchers and advocacy groups for action on rampant online Islamophobia continue to fall on deaf ears. Indeed, the ICV notes that Twitter is still yet to issue a response to our study and has furthermore ignored almost all reports of defamatory and vilificatory anti-Muslim tweets labelling our work as, inter alia, terrorist propaganda. This is unsurprising given that platforms have

38 Butler (n 2) 9.
40 Butler (n 2) 4.
42 Riga (n 10).
44 Copsey et al (n 43) 11 [5.3], 15 [6.3].
45 See, eg, @shrivastavanita (Twitter, 21 September 2022, 8:44pm AEST) <https://twitter.com/shrivastavanita/status/1572537366392471556>.
no real monetary incentive to invest in improving their moderation systems when that would only be likely to reduce user engagement. In the view of the ICV, the only way to address this structural flaw in the operation of social media platforms is for government to intervene, establishing weighty legal and financial repercussions for repeated, systemic and deliberate failures to act against digital Islamophobia.

46 Reset (n 7) [5]-[6].
4 Systemic Problems Demand Systemic Solutions

The traditional regulatory response to the presence of unlawful and harmful material on social media has been to establish content takedown regimes. This is the approach employed by Germany’s Network Enforcement Act 2017 (‘NetzDG’) and, for the most part, Australia’s Online Safety Act 2021 (‘OSA’). Although undoubtedly well-intentioned, we consider such content-focused regulatory regimes to be a piecemeal attempt at solving a systemic problem. We note, for instance, that the fact that platforms are only required to remove material that has been flagged by a user, coupled with the sheer number of harmful posts made each day, makes it practically impossible to ever ensure that a meaningful quantity of prohibited content ends up being taken down. Indeed, in the case of the NetzDG, William Echikson and Olivia Knodt find that, ‘Contrary to expectations, the law has generated only a trickle, not the feared flood, of takedown requests.’

By assigning responsibility for content moderation to users, takedown regimes present another issue. Namely, that platforms are granted free rein to continue developing systems that, to quote a 2018 internal Facebook presentation, ‘exploit the human brain’s attraction to divisiveness ... in an effort to gain

---

47 Reset, Submission No OSB0203 to Commons Digital, Culture, Media and Sport Sub-Committee on Online Harms and Disinformation, UK Parliament (September 2021) <https://committees.parliament.uk/writtenevidence/39851/pdf/>. See also Woods and Perrin (n 8) 11.
48 Gesetz zur Verbesserung der Rechtsdurchsetzung in sozialen Netzwerken [Act to Improve Enforcement of the Law in Social Networks] (Germany) 1 September 2017, BGBl I, 2017, 3352 [tr author], discussed in Reset (n 47).
50 Reset (n 47).
user attention & increase time on the platform'. This failure to address the underlying causes behind the growth of harmful material essentially makes takedown regimes a bandaid solution, one that can alleviate the symptoms of but never really eliminate online harm.

To effectively regulate social media, it is necessary to reform the design of the systems and processes that have enabled and at times encouraged the widespread and unchecked dissemination of harmful content. While we commend the Government for the steps it has already taken towards that end, particularly in introducing co-regulatory codes as well as setting Basic Online Safety Expectations (the ‘BOSE’) for platforms, which require them to, *inter alia*, report to the eSafety Commissioner on how they keep users safe, the ICV does not consider such responses to be wholly adequate for curbing the rampancy of digital Islamophobia. With regard to the BOSE, we note that the only real consequence of a failure to meet an eSafety expectation is the publication of a statement by the Commissioner to that effect. In the case of co-regulatory codes, we share the sentiments of Reset Australia as expressed in their submission to the *Inquiry into Social Media and Online Safety*:

>[C]o-regulatory codes and guidance from regulators will not be adequate to create the scale of change needed to ensure safety. These risks are simply too important to leave up to industry to address — whose business models incentivise and reward risky systems. Nor will the proposed Codes cover all of the systems and processes that need to be addressed. A more comprehensive approach is needed to ensure the regulatory framework is fit for newly emerged and emerging technologies.

The particular regulatory approach preferred by the ICV is to impose an eSafety duty of care. As will become evident in the succeeding section, in holding platforms to legal account for failures to design systems that prioritise user safety over engagement, this regulatory regime constitutes a truly systemic solution to the systemic problem of online harm.

---

52 Horwitz and Seetharaman (n 6)
53 OSA (n 49) pt 9 div 7, ss 46(1)(a), 49(2)(a), 56(2)(a).
54 Ibid s 48(2)(c).
55 Reset Australia (n 49) 2–3.
5 With Great Power Comes an eSafety Duty of Care

The concept of a duty of care originates in the law of negligence, wherein, for a claim to succeed, the plaintiff must show that:

1. The defendant owed them a duty to exercise reasonable care;
2. The defendant breached that duty;
3. The defendant’s breach caused them to suffer damage; and
4. The damage suffered was not too remote a consequence of the breach.\(^{56}\)

Although originally imposed by the courts, duties of care may now also arise under statute, particularly in situations where the common law appears inadequate,\(^{57}\) with a relevant example being Section 19(1) of the *Work, Health and Safety Act 2011* (Cth):

A person conducting a business or undertaking must ensure, so far as is reasonably practicable, the health and safety of:

(a) workers engaged, or caused to be engaged by the person; and

(b) workers whose activities in carrying out work are influenced or directed by the person;

while the workers are at work in the business or undertaking.\(^{58}\)

It is in the context of legislators establishing new duties of care to fill in gaps left by the common law that the idea of an eSafety duty of care emerges.\(^{59}\) Essentially, statute would be introduced to compel platforms to take reasonable care to protect their users.\(^{60}\) The conceiveers of the eSafety duty of care, Lorna Woods and William Perrin, propose the following provision as a starting point: ‘It shall be the duty of every qualifying operator: to ensure, so far as is reasonably practicable, that the users of their service are free from

\(^{56}\) Tame v New South Wales (2002) 211 CLR 317, 348–9 [88].

\(^{57}\) Woods and Perrin (n 8) 29.

\(^{58}\) *Work, Health and Safety Act 2011* (Cth) s 19(1) (‘WHS Act’).

\(^{59}\) Woods and Perrin (n 8) 29–30.

\(^{60}\) Ibid 30.
harm arising from its operation or use.\textsuperscript{61} Under this overarching duty would sit a number of more specific obligations designed to ensure user safety.\textsuperscript{62} Codes of practice would also be developed by an independent regulator through community and industry consultation, allowing platforms to both safely and confidently discharge their duties of care.\textsuperscript{63}

While the UK’s Online Safety Bill 2021 (the ‘Bill’) does not actually impose a single, overarching duty of care to protect users from harm, it does impose a number of other eSafety obligations that make it a particularly useful example of what our proposed regulatory framework might look like in practice.\textsuperscript{64} It is also helpful in that it has been subjected to intense public scrutiny,\textsuperscript{65} allowing for the identification of weaknesses for Australian regulators to improve upon. We note, for instance, that the Bill creates an excessively complex if not confusing system of duties that are contingent not only on the size of a platform but also on whether the content in question is illegal, harmful to children or harmful to adults.\textsuperscript{66} At the same time, the Bill fails to adequately define what is meant by ‘harm’, with the power to determine a set of recognised categories of harmful content furthermore being left to the Secretary of State.\textsuperscript{67} These flaws have led the Bill to be criticised by, on the one hand, researchers concerned that the ill-defined nature of ‘harm’ could be abused to infringe on users’ freedom of expression,\textsuperscript{68} and, on the other hand, advocacy groups fearful that the distinctions made between small and large platforms, and material that is illegal, harmful to children and harmful to adults could lead to certain classes of online harm such as hate speech falling through the

\textsuperscript{61} Ibid.
\textsuperscript{62} Ibid. Cf WHS Act (n 58) pt 2 div 3.
\textsuperscript{63} Lorna Woods and William Perrin, ‘Obliging Platforms to Accept a Duty of Care’ in Martin Moore and Damian Tambini (eds), Regulating Big Tech: Policy Responses to Digital Dominance (Oxford University Press, 2022) 93, 102.
\textsuperscript{64} UK OSB (n 9) pt 3 ch 2.
cracks,\textsuperscript{69} in addition to platforms with small but nonetheless extremely toxic user bases being overlooked.\textsuperscript{70} In recognition of these concerns, the ICV thus recommends the imposition of a duty to take reasonable care to protect users from, in particular, a set of designated categories of harm that are enshrined in statute and thereby subject to parliamentary scrutiny. Furthermore, we propose that this overarching duty of care be filled in by a number of more specific eSafety obligations that apply to all platforms but with varying force depending on, \textit{inter alia}, their size, the toxicity of their user base, the likelihood of harm, the severity of harm and the resources they have available to address harm.

Borrowing from the Bill, the first eSafety obligation we suggest imposing on platforms is a duty to conduct and maintain sufficient, suitable and timely risk assessments that assess the risk of harm and potential harm posed by the functionalities, user bases, business models, governance and other systems and processes of their services to users in general as well as members of vulnerable groups such as children and people of colour.\textsuperscript{71} To enforce this obligation, we envision an independent regulator (the ‘Regulator’), perhaps the eSafety Commissioner or Australian Communications and Media Authority, being empowered to issue guidance on how to carry out risk assessments, set risk profiles and inspect and audit platforms.\textsuperscript{72} Although not provided for by the Bill,\textsuperscript{73} we also find it important that academics and advocacy groups like the Islamic Council of Victoria be granted access to relevant research data that would enable them to conduct their own surveys of online harm, effectively providing a degree of public scrutiny and oversight to the eSafety duty of care regulatory framework.\textsuperscript{74}

Once a risk assessment has been completed, we recommend that a duty should arise to take proportionate measures to mitigate and manage the risks

\textsuperscript{69} See, eg, Ben Whitham, ‘A Focus Too Narrow and an Exemption Too Broad’ in Community Policy Forum (ed), \textit{The Online Safety Bill: Will It Make Online Spaces Safer for Muslim Communities?} (Report, 31 May 2022) 8, 8.\textsuperscript{70} See, eg, Antisemitism Policy Trust, Submission No OSB23 to Commons Online Safety Bill Committee, UK Parliament (25 May 2022).\textsuperscript{71} UK OSB (n 9) cls 8, 10, 12, 190(5).\textsuperscript{72} Ibid pt 7 chs 3–4.\textsuperscript{73} But see ibid cls 89, 137.\textsuperscript{74} Reset (n 7) [11]; Reset Australia (n 49) 25; Center for Countering Digital Hate, Submission No OSB60 to Commons Online Safety Bill Committee, UK Parliament (8 June 2022) 5 [20].
of harm and potential harm identified.\textsuperscript{75} What this means in practice is that, if a platform such as Facebook were to become aware through its conduction of a risk assessment that its algorithms for deciding what content to recommend to users were essentially ‘exploit[ing] the human brain’s attraction to divisiveness’,\textsuperscript{76} thereby causing considerable, unreasonable harm to vulnerable groups targeted by online hate, then they would be legally required to rewrite those algorithms to better prioritise user safety over engagement. At the same time, however, the fact that this duty requires platforms to use ‘proportionate measures’ means that smaller services with less toxic user bases would not be bound to take the same amount of action that a large, well-resourced platform such as Facebook might.\textsuperscript{77} Furthermore, reference to ‘proportionality’ also imports considerations of the likelihood, extent and severity of harm, thus ensuring that platforms are not compelled to unduly restrain users’ rights and freedoms.\textsuperscript{78}

We recommend that the duty of platforms to act on the findings of risk assessments be joined with an obligation to develop clear and accessible terms of service that explain what reasonable steps they are taking to protect users from harm.\textsuperscript{79} These terms should be applied consistently, with mechanisms being created for users to file complaints as well as report harmful content.\textsuperscript{80} This would effectively make platforms legally accountable to their users, granting users the corresponding right to have their service agreements enforced fairly.

Although we believe that platforms should be free to devise their own unique measures for protecting users, we also see benefit in introducing codes of practice for designing systems to better prioritise user safety. These codes could be developed by the Regulator through an evidence-based process that involves consultation with community and industry representatives.\textsuperscript{81} Once a code of practice has been adopted by a platform, they should then be considered to have discharged their eSafety duty of care.\textsuperscript{82} There is, however, one aspect of the regulatory regime established by the Bill for the development of codes of practice that we do not endorse, namely the granting of extensive

\textsuperscript{75} UK OSB (n 9) cls 9(2), 190(5).
\textsuperscript{76} Horwitz and Seetharaman (n 6).
\textsuperscript{77} Woods and Perrin (n 8) 46.
\textsuperscript{78} Ibid.
\textsuperscript{79} UK OSB (n 9) cls 9(5), (8).
\textsuperscript{80} Ibid cls 9(6), 17(2), 18(2).
\textsuperscript{81} Woods and Perrin (n 8) 35.
\textsuperscript{82} UK OSB (n 9) cl 45(1).
powers to the Secretary of State for overriding the decisions of the Regulator. As countless civil rights groups and community advocates alike have observed, there is no clear reason why the executive should be so involved in the day-to-day work of an independent regulator that they can rewrite guidance materials, codes of practice and enforcement strategies, particularly where fundamental human rights are directly at stake.

The establishment of an eSafety duty of care represents one of the best proposed regulatory responses to the pervasiveness of online anti-Muslim hate to date. Unlike a takedown regime, responsibility for the protection of users would be placed solely in the hands of platforms, reflective of the fact that they have an enormous degree of power over their users and so ought reasonably to exercise that power with care. Furthermore, the scope of platforms’ duties to users would extend beyond just content to every aspect of their services, including their functionalities, user bases, business models, governance and other systems and processes. Correspondingly, users would be granted a right to have their agreements with platforms enforced consistently and fairly. In the view of the ICV, it is this focus on the rights of users and responsibilities of platforms that makes the eSafety duty of care a truly systemic solution, one with the potential to seriously stunt the otherwise unchecked growth of digital Islamophobia.

---

83 Ibid cls 40(1), 146-7, 173, sch 4 para 7, discussed in Carnegie UK Trust (n 67) [31]-[32], [34].
84 See, eg, Carnegie UK Trust (n 67) [30]-[34]; Antisemitism Policy Trust (n 70); 5Rights Foundation, Submission No OSH0019 to Commons Digital, Culture, Media and Sport Sub-Committee on Online Harms and Disinformation, UK Parliament, Inquiry into Online Safety and Online Harms (22 September 2021) 13 <https://committees.parliament.uk/writtenevidence/38530/pdf/>.
6 Conclusion

For close to a decade, academics and advocacy groups alike have warned platforms of the rise of online anti-Muslim hate to no avail. Indeed, our research shows that just 14.83% of Islamophobic posts on Twitter end up being removed,\textsuperscript{85} with similar findings being made by the Center for Countering Digital Hate in respect of other major platforms such as Instagram, YouTube and Facebook.\textsuperscript{86} Furthermore, there are also indications that there are tens if not hundreds of millions of new anti-Muslim posts being made each year. At the same time, whether due to overly restrictive definitions of online harm or the simple fact that users cannot be relied upon to identify a meaningful quantity of Islamophobic posts, takedown regimes have proven themselves wholly inadequate for the prevention of digital anti-Muslim hate. A new solution is necessary, one that looks beyond the symptoms of online Islamophobia towards its root causes, namely the systems and processes platforms design to ‘exploit the human brain’s attraction to divisiveness’ at the cost of user safety.\textsuperscript{87} Although there exist a number of proposed regulatory regimes for the implementation of this systemic approach, our preference is to place a statutory duty of care on social media platforms to protect users from harm, thereby shifting responsibility away from users to the real drivers of digital hate. Regardless of the particular regulatory response taken, however, this report should by now have made clear that something must be done. If not to improve the mental wellbeing of Muslim users, then at least to ensure that the events of the Christchurch attacks are never repeated.

\textsuperscript{85} Butler (n 2) 8.
\textsuperscript{86} Center for Countering Digital Hate (n 21) 10.
\textsuperscript{87} Horwitz and Seetharaman (n 6).

Awaad, Rania et al, ‘Suicide Attempts of Muslims Compared with Other Religious Groups in the US’ (2021) 78(9) JAMA Psychiatry 1041

Awan, Imran, ‘Islamophobia and Twitter: A Typology of Online Hate Against Muslims on Social Media’ (2014) 6(2) Policy and Internet 133

Awan, Imran and Irene Zempi, ‘The Affinity Between Online and Offline Anti-Muslim Hate Crime: Dynamics and Impacts’ (2016) 27 Aggression and Violent Behavior 1


Center for Countering Digital Hate, Failure to Protect: Social Media Platforms Are Failing to Act on Anti-Muslim Hate (Report, 28 April 2022) <https://counterhate.com/research/anti-muslim-hate/>

Center for Countering Digital Hate, Submission No OSB60 to Commons Online Safety Bill Committee, UK Parliament (8 June 2022) <https://bills.parliament.uk/publications/46683/documents/1891>


Eckert, Stine et al, ‘*A Hyper Differential Counterpublic: Muslim Social Media Users and Islamophobia During the 2016 US Presidential Election*’ (2021) 23(1) *New Media and Society* 78


5Rights Foundation, Submission No OSH0019 to Commons Digital, Culture, Media and Sport Sub-Committee on Online Harms and Disinformation, UK Parliament, *Inquiry into Online Safety and Online Harms* (22 September 2021) <https://committees.parliament.uk/writtenevidence/38530/pdf/>

*Gesetz zur Verbesserung der Rechtsdurchsetzung in sozialen Netzwerken* [Act to Improve Enforcement of the Law in Social Networks] (Germany) 1 September 2017, BGBl I, 2017, 3352


*Online Safety Act 2021* (Cth)

Online Safety Bill 2021 (UK)


Reset, Submission No OSB0203 to Commons Digital, Culture, Media and Sport Sub-Committee on Online Harms and Disinformation, UK Parliament (September 2021) <https://committees.parliament.uk/writtenevidence/39851/pdf/>

Reset, Submission No OSH0024 to Commons Digital, Culture, Media and Sport Sub-Committee on Online Harms and Disinformation, UK Parliament (September 2021) <https://committees.parliament.uk/writtenevidence/38548/pdf/>


@shrivastavanita (Twitter, 21 September 2022, 8:44pm AEST) <https://twitter.com/shrivastavanita/status/1572537366392471556>

Tame v New South Wales (2002) 211 CLR 317


Work, Health and Safety Act 2011 (Cth)

Zempi, Irene and Imran Awan, Islamophobia: Lived Experiences of Online and Offline Victimisation (Policy Press, 2016)