

# In Defence of Diversity:

A proposal to counter online echo chambers that socialise people towards hate crime and extremist violence.

By R. Jabri-Markwell, Law and policy lead at the Australian Muslim Advocacy Network  
Last updated 5 March 2021



## Table of Contents

<b>INTRODUCTION .....</b>	<b>3</b>
<b>THE ISSUE OF DEHUMANISATION .....</b>	<b>3</b>
<b>OUR OBSERVATIONS OF FACEBOOK AND TWITTER.....</b>	<b>3</b>
<b>EXISTING RESEARCH ON THIS PROBLEM.....</b>	<b>4</b>
<b>HOW THESE MATERIALS AVOID PLATFORM DETECTION OR PENALTY .....</b>	<b>4</b>
<b>EXAMPLES FROM THE ANTI-MUSLIM CONTEXT .....</b>	<b>5</b>
<b>CHOOSING THE APPROPRIATE LEGAL LEVERS .....</b>	<b>6</b>
<b>RECOMMENDATION.....</b>	<b>10</b>
<b>ANNEXURE A .....</b>	<b>11</b>
POSSIBLE DEFINITION FOR “DEHUMANISING MATERIAL” WITHIN THE ONLINE SAFETY ACT .....	11
<b>ANNEXURE B .....</b>	<b>13</b>
INDUSTRY STANDARD OR FRAMEWORK FOR DETERMINING WHETHER AN ACTOR HAS <i>OVER TIME</i> DEHUMANISED A GROUP OF PERSONS IDENTIFIED ON THE BASIS OF A PROTECTED CHARACTERISTIC.....	13
<b>ANNEXURE C.....</b>	<b>15</b>
SAMPLE OF USER RESPONSES (ECHO CHAMBER DISCUSSION) NOT DETECTED BY AUTO-DETECTION ON FACEBOOK AND TWITTER.....	15

## Introduction

1. This proposal is to Government and platforms operating in Australia, but is possibly applicable to jurisdictions abroad.
2. We are sharing this proposal to test our understanding of the issue, to seek further information from other communities, industry and researchers, and to build more consensus about solutions moving forward.

## The issue of dehumanisation

3. Dehumanisation works to help a person overcome normal moral objections they may have to enacting violence against another person or group. The target group is placed in a subhuman or inhuman category and constructed as an existential threat – thus, violence against them becomes proper, necessary, even righteous.<sup>i</sup>
4. From our observations, dehumanisation is usually enacted in insidious ways that circumvent platform policies on hate speech. While much of it appears to skirt beneath the threshold for vilification or incitement if examined post by post<sup>ii</sup>, over time, it creates serious aggregate harm by socialising individuals towards the violent denial of that group's right to co-exist peacefully.<sup>iii</sup>

## Our observations of Facebook and Twitter

5. Our research has found explicitly dehumanising language ('invaders,' 'disease,' 'savages') directed at Muslims is frequently not detected by Facebook's and Twitter's tools.
6. Most of the vilification, incitement to violence and glorification of genocide we observed in comment threads within 'echo chamber' environments were undetected by autodetection frameworks of Facebook and Twitter. This is a common constraint of autodetection frameworks. Examples are provided at **Annexure C**.
7. We found that platform attempts to weed out hate speech and incitement to violence are occurring too late when the targeted group has already been dehumanised in the reader's mind by dehumanising materials.
8. Users within an echo chamber are often responding to materials that seek to dehumanise an outgroup to the in-group audience. The materials, constituted

mainly of links to stories on third party websites, did not trigger platforms' consequences.<sup>iv</sup>

9. In August 2020, we reported 30 public pages to Facebook that were routinely sharing the material of three actors. Facebook's investigation of these pages, including the process and criteria used was ambiguous. Only one page was removed. Previously, when we sought national media attention with evidence about the series of violations in pages and groups, Facebook has acted immediately.
10. At the same time, we reported several accounts to Twitter, to which no action has been taken, although Twitter has advised it is under ongoing consideration. There seems to be some early awareness about aggregate harm and how its current policies are not equipped to identify it.
11. It appears the only way we can currently get Facebook and Twitter to act is to document extensive evidence of violations within the comment threads on every new or existing account, page or group, to make an argument that the account or page admin is failing to moderate. This is beyond our resources and psychologically harmful to the community to continually peruse this material.
12. To support industry to escalate and assess these echo chamber environments, we have engaged research into these actors to extract predictors that could form part of a universal assessment tool.

## Existing research on this problem

13. The problematic materials we have identified are part of the 'counter jihad movement.' According to scholars like Benjamin Lee, Melagrou-Hitchens and Brun, and others, the 'counter-jihad' movement is classified as an extreme-right movement.<sup>v</sup> Unlike some extreme-right movements, however, the counter-jihad movement tends to avoid placing itself firmly within the white supremacist space by engaging in superficially liberal critiques of Islam, all while maintaining a steady diet of anti-Muslim stories. By giving false ideological context to contemporary events, Muslims are a 'hostile and homogenised mass' that seeks to overtake the West.<sup>vi</sup>
14. The practice of sharing disinformation to vilify or dehumanise an identified group over time will be an issue to other segments of the community.. However, an online scan would be needed to analyse permeations in a range of contexts.

## How these materials avoid platform detection or penalty

15. Mainstream platforms like Facebook and Twitter have dehumanisation policies, but they focus on explicitly dehumanising language, a feature that 'auto-detection' systems can detect. However, our studies have found the presence of

blatantly dehumanising terms is not necessary to effectively dehumanise 'the other'.

16. Actors conveyed dehumanising conceptions by the headlines and content of 'stories' published to Facebook and Twitter via third party link.

## Examples from the Anti-Muslim context

17. It would appear that Facebook and Twitter are *still unclear* on whether conceptions from 'counter jihad' ideology – including propagating that (1) personal religiosity in Islam in itself leads to sub-humanity and extremism, and (2) Islam/Muslims are invading the West to take over through immigration and higher fertility rates, and (3) Islam/Muslims are in a clash-of- civilisations style violent 'jihad' war with the West – are harmful and dehumanise Muslims.
18. The latter narrative is also part of ISIS-inspired propaganda, showing the two directions in which this narrative is coming. But while ISIS propaganda is treated as violent and extremist content, the Western extreme right's propaganda is not.
19. Demographic invasion and replacement theories about Islam and Muslims are grounded in dehumanising conceptions of Muslims, as evidenced by the responses they illicit. This included the portrayal of Muslims as:
  - a. mechanically inhuman<sup>46</sup> 'theological automatons' who are 'unified in thought and deed' to carry out demographic invasion.<sup>47</sup> Significantly, it follows that there is no way to tell if Muslims are truly peaceable or not, and therefore all Muslims are a threat.
  - b. Subhuman in their inherent violence, barbarism, savagery, or in their plan to infiltrate, flood, reproduce and replace (like disease, vermin without explicitly using those terms).
20. We have identified Islam as a proxy for Muslims in 'counter jihad' contexts, which is uncovered through the language technique of personification. For example:

*'Islam exists in a fundamental and permanent state of war with non-Islamic civilizations, cultures, and individuals (a group of people, not a religion, can be in a state of war with civilisation)'*

*'A halt to terrorism would simply mean a change in Islam's tactics — perhaps indicating a longer-term approach that would allow Muslim immigration and higher birth rates to bring Islam closer to victory before the next round of violence'*

*'Islam proper remains permanently hostile'*

'Islam's *violent nature* must be accepted as given'

In counter jihad context, Islam is attributed human actions and qualities as a seemingly more liberal route to vilify and dehumanise Muslims as both subhuman and mechanically inhuman species.

21. Even labels such as cancer and disease imply that Islam is growing, which again points to Muslims, the religion's followers, as the existential threat. This is revealed by the solutions that users also point to for this cancer or disease, including the deportation, extermination, or forced conversion of Muslims.
22. Our studies also showed that while digital platform may be looking for dehumanising descriptors (*adjectives or synonyms*), dehumanising discourses are also cumulatively and powerfully conveyed in headlines through
  - a. *verbs* associated with the subject 'Muslim' (eg., 'stabs,' 'sets fire') and
  - b. *essentialising the target identity* through implicating a wide net of Muslim identities (eg., 'Niqab-clad Muslima,' 'boat migrants,' 'Muslim professor,' 'Muslim leader', 'Iran-backed jihadis', 'Ilhan Omar') to suggest they are acting in concert.
23. Social media companies may rightly question how to identify whether a 'news story' is merely reporting news or opinion about human rights abuses, foreign affairs or violent extremism, rather than operating as part of a concerted dehumanisation project by a specific actor. This behaviour needs to be analysed with regard to contextual factors, which ought to be distilled in a transparent and explicable way to guide more competent and consistent assessments. The aim is to articulate how even lawful but harmful speech can cause serious harm in aggregate over the long-term in certain contexts.

## Choosing the appropriate legal levers

24. It is a challenge to consider the dispersed social harm that stems from dehumanisation into an individualistic frame. Our recommendations have focused on the vectors of this harm, that being individuals who serially post dehumanising material; and through setting an industry standard for digital platforms when making detailed and contextualised assessments about individual accounts, pages, groups and channels. As civil provisions, this would create a consequence for both individuals serially engaged in this practice, along with platforms that disregard it. As civil provisions, it is also possible to set aside the requirement often put forward in criminal contexts that there be evidence of foreseeable or imminent physical harm.
25. The Rabat Plan also emphasises context: of the speaker's power, their intent, the content and form, spread, and likelihood and imminence of harm. While imminence of harm would not be a necessary threshold requirement for the civil penalty we have proposed, the other contextual factors would be

considered. It is also vital that targeted communities are consulted on their particular contexts as otherwise decision-makers will fail to make fully competent judgements.

26. The Rabat Plan of Action noted the importance of distinguishing not just criminal and civil prohibitions, but on a broader class that will “still raise concerns in terms of tolerance, civility and respect for the convictions of others.” If we limit civil prohibitions to the most severe end of spectrum (serial and clear-cut examples) and invoke the Act’s Basic Online Safety Expectations and an Industry Standard as levers to engender platform accountability on a broader range of dehumanising speech or discourse, this will go a long way to satisfy Australia’s obligations under international human rights law in terms of protecting freedom of expression.
27. As the Global Internet Forum to Counter Terrorism has recently acknowledged, platforms tend to focus on ISIS and Al-Qaeda inspired propaganda because legal frameworks define extremist or terrorist content in line with official proscription lists, and rely heavily on the identification of organisational symbols.
28. The bias towards a taxonomy that relies on external designation lists ignores the contemporary reality of radicalisation and recruitment materials. For example, online echo chamber forums that socialise individuals towards violence, typically include many examples of the materials that initiatives like the GIFCT is trying to prevent from being shared – but are rarely detected by platforms as they are buried within comment threads and lack organisational labels.
29. Proposals to expand designation and proscription lists have struggled with the political and legal difficulty of defining ‘extremist ideology’ or ‘extremist rhetoric’ where there are no explicit or imminent calls to violence. The scope for ‘terror-scaping’ ideas, organisations or individuals, merely because they present as extreme, unpopular or fringe, is a real concern, especially for marginalised communities that are already subject to overpolicing, and may have legitimate grievances with nation-states.
30. Therefore it is imperative to develop a capability for assessing materials that have a close nexus to socialising individuals towards extremist violence, in a way that is ideology-agnostic and clearly defined.
31. Previous attempts of policy making in this area tend to oscillate between very general approaches (eg UK’s failed Bill to ban extremist speech in 2015<sup>vii</sup>) and specific approaches, often adopted by platforms, to list the types of hate speech or incitement that will not be accepted. This means that organisations or websites that serially attempt to socialise individuals towards extremist violence are missed, especially when they skirt beneath the threshold of hate speech or criminal incitement<sup>viii</sup> (for example through disinformation).
32. It is noted that in the UK, a recent review<sup>ix</sup> by the independent Commission for Countering Extremism has recommended establishing a legal framework to counter hateful extremism, which it has defined as:

activity or material directed at an out-group” (e.g., Muslims) who are perceived as a threat to an in-group (e.g., a Far-Right group) “motivated by or intending to advance a political, religious or racial supremacist ideology: a. To create a climate conducive to hate crime, terrorism or other violence; or b. Attempt to erode or destroy the fundamental rights and freedoms of our democratic society as protected under Article 17 of Schedule 1 to the Human Rights Act 1998 (‘HRA’).

33. Their report emphasises that this is a working definition, not a legal one. It also emphasises that hateful extremism should be treated with as much priority as terrorism.
34. Defining extremist material or activity at law is more fraught, especially in terms of articulating a clear actus reus (the act, as opposed to the intent). This ambiguity can create anxiety about big state or big tech interference in freedom of speech.
35. The same can be said for vilification legal standards, involving the incitement of hatred, severe ridicule or serious contempt – again descriptions of conduct which rely on assessments about intent and effect. This makes vilification standards a more complex tool for implementation by administrators under a civil penalties scheme.
36. Dehumanisation offers an enduring, internationally accepted<sup>x</sup> and well-defined<sup>xi</sup> concept, grounded in genocide prevention studies<sup>xii</sup>, and increasingly in literature on countering violent extremism.<sup>xiii</sup> Most violent extremist movements tend to rely upon dehumanisation of an ‘out-group’ to their ‘in-group’ audience.
37. Given the difficulties in determining the bounds of ‘extremist material’ as per the Criminal Code, proscribing dehumanising materials through the Online Safety Act is a way of taking action on conspiracy theory propaganda without intruding upon legitimate speech that is otherwise regarded as extreme, unpopular or fringe.
38. Dehumanisation is carried out through language and discourse, portraying the target group as<sup>xiv</sup>:

<b>Subhuman</b>	<b>Mechanically inhuman</b>	<b>Supernatural</b>
The material presents the class of persons to have the appearance, qualities or behaviour of an animal, insect, form of disease or bacteria; or the material suggests that the whole class of persons are polluting,	The material presents the class of persons be inanimate or mechanical objects; or the material suggests the class of persons acts in concert to harm the in-group and are	The material presents the class of persons to be supernatural threat.

despoiling or debilitating society.	incapable of human thought or feeling.	
--	---	--

39. Much dehumanisation occurs in a gradual and cumulative way, through disinformation and conspiracy narratives.
40. While it may be tempting to set the threshold higher at incitement to violence, incitement to violence is
- (a) a difficult concept given platforms (and criminal contexts) demand it pose an imminent threat – creating a slippery and impractical evidentiary burden.
  - (b) Additionally, a person who uses violent language is often reacting to dehumanising materials directed to them as an in-group audience member. Acting on incitement to violence is acting too late.
  - (c) The most prevalent and harmful forms of weaponization of digital platforms is not by organisations or websites openly inciting, threatening or glorifying violence, but inducing and inspiring it, through dehumanising materials about outgroups to in-group audiences.

## RECOMMENDATION

Make **dehumanising material** unlawful and harmful content in Australia's Online Safety Act (OSA). There are several steps to doing this:

- (1) Define dehumanising material in the Act. Wording has been proposed in **Annexure A** to start discussion.
- (2) Include an additional and distinct civil penalty in relation to the *serial publication of dehumanising material*. Initially, potential targets for this penalty include:
  - a. Groups that in the aggregate are spreading large amounts of serious negative dehumanising discourse (also referred to as 'echo chambers')
  - b. Individuals who are intentionally carrying on a campaign
  - c. Platforms that implicitly or explicitly allow either of the previous
- (3) Australia's e-Safety Commissioner creates an industry standard regarding the assessment framework platforms ought to use to consistently and competently identify an individual or individuals who are engaged in dehumanisation over time (creating an aggregate harm).

This would include the serial publication of 'stories' where a nexus with dehumanisation can be established. Our research has analysed the actor behaviour with regard to contextual factors, which is distilled into a proposed industry standard at **Annexure B**.

## ANNEXURE A

### Possible definition for “dehumanising material” within the Online Safety Act

- (1) This section sets out the circumstances in which material is dehumanising of a class of persons for the purposes of this Act.

#### *Dehumanising language*

- (2) Material is dehumanising of a class of persons if:
- (a) The material presents the class of persons to have the appearance, qualities or behaviour of an animal, insect, form of disease or bacteria; or
  - (b) The material presents the class of persons to be inanimate or mechanical objects, which are incapable of human thought or feeling; or
  - (c) The material presents the class of persons to be supernatural threat

In circumstances in which a reasonable person would conclude that the material was intended to cause others to see that class of persons as less deserving of being protected from harm or violence.

#### *Implicitly dehumanising disinformation or discourse*

- (3) Material is dehumanising of a class of persons if:
- (a) The material presents that evidence of a person committing a heinous crime is proof that this person’s entire group, on the basis of a protected characteristic, has subhuman qualities; or
  - (b) The material presents that a class of persons are to be held responsible for, and deserving of collective punishment for the specific crimes, or alleged crimes of some of their ‘members’; or
  - (c) The material expresses that the whole class of persons are polluting, despoiling or debilitating<sup>xv</sup> society

In circumstances in which a reasonable person would conclude that the material was intended to cause others to see that class of persons as less deserving of being protected from harm or violence.

- (4) Context may be considered to determine if the conditions in subsections (2) and (3) have been satisfied, including the
  - a. Form of the material
  - b. Speaker's power or influence
  - c. Audience responses to the material
  - d. Forum or forums where it is posted
  - e. The content contained on a website or social media page that is publicly linked to a forum where the material is shared
- (5) It is not necessary to establish the risk or imminence of physical harm to satisfy conditions in subsections (2) and (3).
- (6) *Class of persons* means a group identified on the basis of a protected characteristic, such as religion, ethnicity, nationality, race, colour, descent, gender or other identity factor.<sup>xvi</sup>
- (7) Dehumanising material that is not directed at a protected characteristic is not included.

## ANNEXURE B

Industry Standard or Framework for determining whether an actor has *over time* dehumanised a group of persons identified on the basis of a protected characteristic.

xvii

The following predictors could be used to assess aggregate conduct that has dehumanised an identified group:

1. *Dehumanising conceptions on the actor's website in relation to an identified group.* This may be expressed explicitly on the website through language or narratives that portray the identified group as subhuman<sup>xviii</sup>, mechanically inhuman<sup>xix</sup> or supernaturally inhuman<sup>xx,xxi</sup>
2. *The features of material that are serially published, specifically*
  - i) *The subjects or participants routinely identified in material.* Analysts will be looking for signs of essentialising an identity as part of a dehumanising discourse about an 'outgroup'. For example, their identity (eg Muslim, Jew, trans, Aboriginal, Black) is routinely emphasised in material to collectively attribute guilt for a specific member's heinous crimes, or to suggest over time that all members of that group act in concert.
  - ii) *Hostile verbs or actions (eg stabs, sets fire) attributed to those subjects to cumulatively associate them with sub- humanity, barbarism, or serious threat to the in-group.*
  - iii) *Use of explicitly dehumanising descriptive language (eg cancer, disease, species, frothing-at-the-mouth, snakes) or coded extremist movement language with dehumanising meaning (eg invader, a term used in RWE propaganda to refer to Muslims as a mechanically inhuman and barbaric force).*
  - iv) *Proportion of actor's material that act as 'factual proofs' to particular narratives about this identified group.* Here, narratives could be defined as narratives that have been used previously to justify atrocities or violence against this identified group.
  - v) *Presence of 'baiting' content to in-group audience.* Rhetorical techniques like irony to draw an even more hateful response towards the identified group.
3. *Evidence in the user comment threads of a pattern of hate speech against a group on the basis of a protected characteristic.* This would include blatantly dehumanising remarks, iteration of extremist ideology concerning the target group as an existential threat to the in-group, or glorification of, or incitement towards, violence against the target group. Where that pattern is evident in relation to a high proportion of links shared from one host website, this can be taken as a primary sign that the website is engaged in a project of hatred or

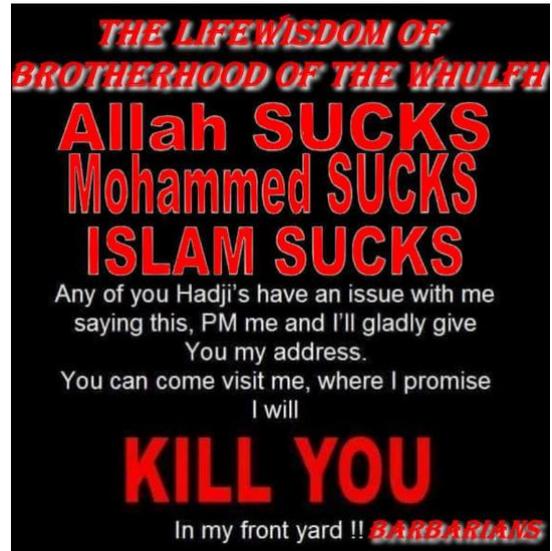
dehumanisation. However, the absence of comments does not signify that dehumanisation has not been successfully enacted in the user's mind.

Sample of user responses (echo chamber discussion) not detected by auto-detection on Facebook and Twitter

- (1) Audience responses to the article 'Paris update: Muslim beheaded teacher in street because he had shown Muhammad cartoons in class', shared on this actor's Twitter account, included dehumanising references to Muslims (separate to the murderer) as a cancer, virus, animals, and savages, and spawned significant commentary on the 'existential crisis' faced by France and the Western world from Islamic invasion, aided by liberals and the political establishment (with exception of Trump).

Audience responses to the same article on Facebook also revealed how these captured audiences interpret acts of terrorism and extremism conducted by ideologically motivated Muslims, and the frequent tendency to attribute blame to all Muslims and Islam, rather than the perpetrators alone. However, in this example on Facebook, it also escalated quickly to fantasies about violence. On Actor A's Facebook page, users responded with *dehumanising insults* ('They are worse than rabid animals, no brains of their own and vile to the core,' 'MOSLEMS ARE INCOMPATIBLE WITH HUMANKIND,' 'never trust them they are two faced. Like two people in one being. Ultimately their loyalty is towards Islam which is evil. If they never change their views on Islam no Matter how friendly, caring, compassionate they seem. If it came down to it they can become the most evil vile & depraved creature'); *calls to expunge Muslims* ('Do not let this atrocity happen in the US, vote the squad out, they are the enemy of mankind'); *repetition of demographic invasion/white genocide theory* ('They don't come to assimilate into western society, they come to dominate and conquer the infidels!! Wake up sheeple, these are barbarians!!', 'The ppl of Europe have to be detoxified from the twin evils of multiculturalism and diversity and then get rid of the leaders that spew lies and willingly put their own citizens to danger and evil'); *glorification of genocide of Muslims* ('The muslims are the only people on Earth who will earn their genocide, but they will be the only genocided people for whom nobody will have a drop of tear'); *calls to war* ('Europe has been Invaded and occupied by Muslims, who have claimed Europe as theirs, since they have Proclaimed Sharia Law!

NATO will have to declare War on the European Islamic Caliphate and Attack European Muslim Strongholds, if they want to become an Independent Europe again?', 'this cult should have its head cut off before it is too late ,have you ever thought about when the oil runs out this cult will be looking at us ,and they will show no mercy'); and *calls to vigilante violence* 'Servicemen only ask: CAN WE GO KILL THESE FUCKERS YET ..... Barbarians/E.F.Whulfh' posted by a user along with the meme to the right)



In one Australian Facebook page that routinely shares this actor's articles, the users responded to this article about the Paris beheading with: 'Go in hot an shoot the lot' (which attracted 7 'like' and 'love' reactions), 'U let them in, they multiply rapidly n impose their will on you. High time France takes the upper hand. Learn from China n Russia.' and 'Time to behead all paedophile moslems. NOW....'.

- (2) A story headline from the same actor, 'Muslims migrate to Australia, file complaint with Human Rights Commission because food they're given isn't halal' produced numerous responses expounding on demographic invasion and replacement. Common dehumanising conceptions from those on Twitter were that Muslims originate from 'cesspools,' 'toilet bowl countries', and 'shitholes', and that resisting their plot had to be done for the sake of 'civilised world and culture.' It appeared to 'trigger' users who saw this as an attempt to 'placate the Moslem invaders'. One user commented, 'Physical appearance of mooslems is like normal human being but mentally like cold blooded demon, Ogre.' The world 'infiltrate' was preferred to migrate. Many spoke about the 'stages' of 'jihad' in taking over a country: 'It starts with halal food, next is burning cities and killing infidels.' While others lamented that the west was contributing to its defeat: 'A secularism & multiculturalism is a breeding ground for deadly peaceful community virus (Islam).' The disgust prompted by this headline also led to calls to expunge: 'What are the options available with Australia? Will they let the cancer spread there also like it has in Europe?'

(3) Another Australian Facebook page, with more than 120 000 followers, routinely shares third party links from an Australian based 'counter-jihad' actor. In 2019 they posted a cartoon meme explaining the premise of the 'great replacement theory.' It compared a Muslim and non-Muslim family in terms of their number of children. The meme was accompanied by similar derogatory statements implying that Muslims plan to conquer countries like Australia through higher fertility rates. The intense reactions to this poster were revealed in the extensive comments, with a significantly high proportion employing dehumanising language, as well as expressions of wanting to kill or see Muslims dead. Some responses included: 'Shoot the fuckers', 'Islam is a cancer on global society for which there is no cure', 'You import the 3rd world you become the 3rd world. And when they become the majority then what next? They won't have whitey to leech off. Just like locusts, infest & strip everything until there is nothing left', 'Deport the PEDO crap', 'They breed like rats', 'Disgusting religion. Ugh On the outside they hide their bodies but under cover they turn into raging sex addicts, breeding faster than rabbits. Tarts in hidden cloth', 'if we get our guns back we can take back parliament and force these idiots out,' 'Drown em at birth', 'Fun those scumbags.muslims....reminds me of aids', 'Society should start culling the Muslims,' 'I think I now understand why during the serbian / croat the serbs culled the women,' 'I'm going out tonight to do as much as i can to solve this problem.'

---

<sup>i</sup> Dehumanisation is a concept recognised in international genocide prevention law and expounded in literature:

- 
- Jonathan Leader Maynard and Susan Benesch, 'Dangerous Speech and Dangerous Ideology: An Integrated Model for Monitoring and Prevention' (2016) 9(3) *Genocide Studies and Prevention: An International Journal* 70.
- Nick Haslam. (2006). Dehumanization: An Integrative Review. *Personality and social psychology review: an official journal of the Society for Personality and Social Psychology*, 257.
- ii Will Baldet, "How 'Dangerous Speech' Is The Mood Music For Non-Violent Extremism: How do we define websites, groups and individuals who stay the right side of our hate crime laws but whistle the tune which advances the rhetoric of violent extremism?", *Huffpost*, 9 May 2018.
- iii Khalifer Ihler Global Institute define violent extremism as the violent denial of diversity: "Unifying all violent extremists, regardless of their beliefs or ideological objectives is their beliefs that peaceful coexistence with someone different from them is impossible, and that violently enforcing this either through forced submission or through eradication of diversity is the solution." Khalifa Ihler Institute, 'Hate Map: Definitions, Scope, Terms', < <https://www.khalifahihler.org/hate-map>>.
- iv Examples of these materials and user reactions are included in a paper that is under publication by AMAN. There are 5 actors we have been observing. One appears to have multiple related websites without any editorial transparency or guidelines.
- v Benjamin Lee, 'A Day in the "Swamp": Understanding Discourse in the Online Counter-Jihad Nebula' (2015) 11(3) *Democracy and Security* 248, 251-3; Alexander Meleagrou-Hitchens and Hans Brun, *A Neo-nationalist Network: The English Defence League and Europe's Counter-jihad Movement* (London, 2013).
- vi Lee, *ibid*, 252.
- vii John Ware, 'Why Britain must not let extremists operate with impunity', *The Article*, 24 February 2021 <https://www.thearticle.com/why-britain-must-not-let-extremists-operate-with-impunity>
- viii See for example, Will Baldet, "How 'Dangerous Speech' Is The Mood Music For Non-Violent Extremism: How do we define websites, groups and individuals who stay the right side of our hate crime laws but whistle the tune which advances the rhetoric of violent extremism?", *Huffpost*, 9 May 2018; James Grierson, 'UK extremists 'exploiting gaps in law to push their agenda'', *The Guardian*, 10 June 2020. <<https://www.theguardian.com/society/2020/jun/10/uk-extremists-exploiting-gaps-in-law-to-push-their-agenda>>; Lizzie Deardon, 'New laws needed to tackle 'shocking and dangerous' scale of extremism, review finds', *The Independent*, February 2021 <https://www.independent.co.uk/news/uk/home-news/extremism-laws-review-impunity-mark-rowley-b1806349.html>
- ix Released 24 February 2021 <https://www.gov.uk/government/publications/operating-with-impunity-legal-review>
- x 'Genocide begins with 'dehumanization;' no single country is immune from risk, warns UN official', *UN News*, 9 December 2014.
- xi Haslam, above n 1.
- xii Maynard and Benesch, above n 1.
- xiii Department of Security Studies and Criminology. (2020, October 9). *Mapping Networks and Narratives of Online Right-Wing Extremists in New South Wales (Version 1.0.1)*. Sydney: Macquarie University. Marczak N. (2018) *A Century Apart: The Genocidal Enslavement of Armenian and Yazidi Women*. In: Connellan M., Fröhlich C. (eds) *A Gendered Lens for Genocide Prevention. Rethinking Political Violence*. Palgrave Macmillan, London.
- xiv These categories draw from both Maynard and Benesch and Haslam's work.
- xv Maynard and Benesch, above n 1, 80.
- xvi Taken from the UN Definition of hate speech: United Nations Strategy and Plan of Action on Hate Speech Detailed Guidance on Implementation for United Nations Field Presences, September 2020, <https://www.un.org/en/genocideprevention/documents/UN%20Strategy%20and%20PoA%20on%20Hate>

---

[%20Speech\\_Guidance%20on%20Addressing%20in%20field.pdf](#). This ought to be considered in context with existing categories of protection in Australia, and include some consultation in regard to most targeted groups. The Australian Hate Crime Network also highlights disability as a targeted group.

- xvii Ibid; The Australian Hate Crime Network generally recommends that the list include identity based on race, religion, gender, gender identity, sexuality or disability.
- xviii The material presents the class of persons to have the appearance, qualities or behaviour of an animal, insect, form of disease or bacteria. Or the material suggests that the whole class of persons are polluting, despoiling or debilitating society (a description used by Maynard and Benesch, above n 1, 80.
- xix The material presents the class of persons be inanimate or mechanical objects; or the material suggests the class of persons acts in concert to harm the in-group and are incapable of human thought or feeling.
- xx The material presents the class of persons to be supernatural threat.
- xxi Where an ideology is not explicitly identified by the site, as the Institute for Strategic Dialogue has done in these circumstances, a sample of the site's produced material could be subjected to qualitative assessment. The other factors listed above would assist in that assessment.

## BIO

Rita Jabri Markwell is a lawyer who studied both law and political science. Her career has spanned two decades of working in federal politics, policy development and advocacy. Most recently, she has been working with the Australian Muslim Advocacy Network to interrogate how propaganda supportive of the Christchurch terror attack and the terrorist manifesto continues to survive on mainstream platforms. This has involved much direct engagement, problem-solving and testing of various ideas with industry (Facebook and Twitter), research and civil society sectors. In 2020, she spearheaded a study of five actors on Facebook and Twitter to understand what predictors ought to be used by platforms to measure aggregate dehumanisation.