



*Submission to the Select Committee on Foreign Interference through Social Media
Department of the Senate, Parliament House, Australia
17 July 2020*

1. The Australian Muslim Advocacy Network (AMAN) is a policy development body working to secure the physical and psychological safety of Australian Muslims. We wish to provide some observations about issues with enforcement of Facebook's Inauthentic Behaviour policy as it relates to your Terms of Reference.

Australian impact

2. In the last federal election, there were approximately 12 fringe parties running with a discriminatory anti-Muslim policy – this is the largest number of groups that we have recorded. We remain very concerned about the exportation of RWE rhetoric from the UK, Europe and USA to Australia through coordinated exercises on social media platforms like Facebook, and its potentially devastating impacts for Australia's democracy, social cohesion and national security.

Chronology of events

3. In December 2019, an investigation by Guardian news revealed an overseas commercial enterprise that was

using its 21-page network to churn out more than 1,000 coordinated faked news posts per week to more than 1 million followers, funnelling audiences to a cluster of 10 ad-heavy websites and milking the traffic for profit.

The posts stoke deep hatred of Islam across the western world and influence politics in Australia, Canada, the UK and the US by amplifying far-right parties such as Australia's One Nation and vilifying Muslim politicians such as the London mayor, Sadiq Khan, and the US congresswoman Ilhan Omar.
4. In early 2020, AMAN and Birchgrove Lawyers conducted an investigation of Facebook's efficacy in enforcing its own hate policy standards. This was on the back of unpublished research by Victoria University in 2018, which studied over 41,000 posts in far-right Facebook groups, identifying radicalising discourse. The most prevalent mobilisation frames included five narratives about Muslims and Islam. We wanted to test whether those groups were still active 2 years later, even after Christchurch, and whether they will still propagating the same radicalising messages.
5. Our investigation of the groups we could find (some had slightly different names) revealed they were still very active. At the time Facebook welcomed our investigation, even though it was reported in national media, and acted on those pages. Facebook has since indicated that they find 'escalation' by civil society to be very useful in identifying harmful pages and groups because their capacity for internal escalation still needs significant work. We

continue to engage with them on policy and enforcement issues, but part of our work is ongoing monitoring to see whether systemic changes are having a translatable impact to make the platform safer, and encourage better moderation by page administrators. So far, we have observed negligible change to their internal escalation capacity.

6. In the course of this work, we have identified what appears to be inauthentic behaviour between a network of pages in Australia, that links to RWE and white supremacist content overseas.

General observations

7. Enforcement by Facebook of its inauthentic behaviour policy appears to be compromised by the following factors:
 - a. Internal mechanisms for escalating possible breaches needs strengthening, forcing them to rely upon external actors to monitor and raise occurrences for investigation.
 - b. Due to a lack of transparency requirements in Australia, it is not possible to ask Facebook to reveal if a Page is run by who it claims to be run by, even if it is in public interest or important to running a civil remedy claim.
 - c. Facebook's tools which promote similar pages and groups may be promoting other pages and groups run by the same admins, which presents an issue when those accounts may be fake or examples of foreign influence.
 - d. We can point out signs of international coordinated behaviour and content-sharing by those involved in 'organised hate', but are reliant on the interest and good will of platforms to investigate.
 - e. Facebook will dedicate resources to uncovering inauthentic behaviour in a region, but this depends on their will and interest to investigate a particular region.
 - f. The interaction of the inauthentic policy with their misinformation and hate speech policies needs consideration. Domestic and foreign actors may be working in concert to spread malicious or false news to foment hatred, but the sharing of links to external sites means they can circumvent hate speech policy, and misinformation is only independently tested by Facebook in a fraction of cases. It remains unclear as to how far Facebook's definition of 'opinion' is protecting external 'news' sites from fact-checking, or from flagging as an unreliable or harmful source.

Facebook Strategies

8. We note recent changes from Facebook to its platform algorithm, such as promoting 'original sources' of news, rather than rehashed news, which we assume is to partially deal with the malicious distortion of news. However, we are yet to discern the impact of this change on the particular network we've observed. Facebook says that it is also trying to inform users of the location of admins of very large pages and groups, so they will know if the admins are based overseas. This information will need to be advertised to users more clearly.
9. Facebook currently relies on the official terror listings of countries, along with its own list of white supremacist organisations (which it developed following Christchurch after realising that government lists were grossly inadequate in that regard). Facebook uses that list to identify 'terror' or 'organised hate' groups by their names, emblems, logos, branding and

slogans. It does not have a content moderation guide for identifying white supremacist ideology or discourse, which may explain why it has delayed/not acted in implementing its own audit's recommendation to ban the praise or promotion of white supremacist ideology on its platform, in the content policy. We also acknowledge that while strengthening its content policy is vital, especially to deal with demonisation and threat construction of minorities, which are currently largely omitted, that enforcement will be challenging, especially given the reliance of organised hate actors on misinformation and malicious/distorted information disguised as external 'news' sites or 'opinion'. This behaviour is relevant to your inquiry because the majority of those sites are based overseas yet promoted to Australian audiences, in what appears to be a coordinated exercise.

External suggestions to empower users

10. We acknowledge external efforts to strengthen user knowledge about the origins and true purpose of material they are reading on social media platforms, such as the 'content authenticity' initiative. This initiative is about creating an industry standard for digital content attribution. The Content Authenticity Initiative is building a system to provide provenance and history for digital media, giving creators a tool to claim authorship and empowering consumers to assess whether what they are seeing is trustworthy (contentauthenticity.org). We are in the early days of researching such recommended measures ourselves.

11. As Evelyn Douek, an Australian expert researcher and commentator based in the US who recently presented at a hearing of your inquiry [said](#):

It may be that we need far more radical reforms than individual CIB-hunting operations (reforms centered on transparency and changing algorithmic amplification) to make sure public discourse isn't exploited and manipulated in corrosive ways.

12. It is noted that the proposed Online Safety Act, which is due to be introduced to Parliament following a consultation process at the start of this year, may offer some benefits in terms of transparency requirements on platforms. The Australian Government has begun by asking industry to develop a voluntary disinformation code, and ACMA has recently released a discussion paper on this. AMAN is looking to consider this in more detail.

Criminal law deterrents

13. Research to date has observed that this radicalising material is often circulated through malicious 'news' platforms to circumvent criminal sanction. This conduct is quite intentional and internationally coordinated, as also shown by Facebook's own work into coordinated inauthentic behaviour. In our investigations, we have observed that almost all of this content falls short of actually inciting violence, which poses a real difficulty in terms of the application of criminal law. Individual actors who are repeatedly using a carriage service to propagate the arguments of Tarrant and Breivik within Australia cannot continue to operate with impunity.

14. This material, when repeatedly posted, can however inspire people to commit or condone mass casualty violence against Muslims (or Jewish people, other minorities) by constructing these groups as an existential threat to the 'in-group'. In the absence of direct and imminent threats of violence, this speech is *at most* captured by some civil remedies, despite its significance as a public harm.

15. One option is to strengthen existing criminal laws that are used for harmful online conduct, such as s474.17 of the federal Criminal Code (using a carriage service to menace, harass, or cause offence). AMAN is consulting on whether a possible amendment is needed. If it is, it could be clarified that this offence applies to conduct which targets a community based on a protected characteristic, if it is intended or is reasonably likely to cause a person from that group to have a reasonable fear for their safety or security of property.

Annexures A and B include more details, which we provide in confidence to your committee (not for publication). Thank you for the opportunity to make this submission.