Internet 2.0 submission to the Senate Finance and Public Administration References Committee Administration of the referendum into an Aboriginal and Torres Strait Islander Voice

23 April 2023

By:

David Robinson

Chair & Co-Chief Executive Officer, Internet 2.0

Strategic Intelligence Advisor to CI-ISAC

Retired Australian Army Intelligence Officer

Level 1, 18 National Circuit, Barton, ACT, 2600, Australia
ABN: 17 632 726 946 (Internet 2.0 Pty Ltd)

Contents

Summary

This submission has been submitted by Internet 2.0 to the Senate Finance and Public Administration References Committee Administration of the referendum into an Aboriginal and Torres Strait Islander Voice. It is the primarily the opinions of the author and Co-CEO Robert Potter.

The response is framed only to the following points in the inquiry's terms of reference: protections against the potential for foreign actors to seek to influence the outcome or public debate on the referendum question; the detection, mitigation, and obstruction of potential dissemination of misinformation and disinformation, including via social media or technology platforms; and the ongoing integrity and assurance processes of the Australian Electoral Commission.

For malign actors that wish to influence Australian elections and referendums there exists real capability for them to conduct large inauthentic social media campaigns that is software enabled. The impacts of these campaigns historically are division in society and loss of trust in the outcomes of elections. In terms of intent, we cannot assess either way if there has or has not been significant attempts in the past to conduct large inauthentic social media campaigns because in our view no serious review has occurred. To review previous elections a large data driven review would need to be conducted.

In our assessment current platforms used by social media platforms to identify these inauthentic social media campaigns are inaccurate. Social media companies also have no long-term incentive to identify and remove inauthentic social media accounts. Technology breakthroughs in generative artificial intelligence (AI) and "deep fakes" are also making it harder over time to accurately detect inauthentic content.

To defend voter confidence in elections and prevent more division within society we recommend multiple pillars to underpin a successful strategy. These are: social media monitoring to accurately identify and assess interference attempts; regulation that forces the removal of inauthentic social media accounts discussing elections and referendums in a fair but fast manner; and awareness and reporting campaigns against foreign interference aimed to inoculate the public before a vote to inauthentic social media accounts.

Capability and intent for hostile actors

There exists real capability for foreign malign actors to seek to influence the outcome of the public debate through the use of inauthentic social media accounts, social media bots controlled by software or hostile states using sophisticated psyops campaigns to conduct foreign influence operations. The scope of this capability is quite significant. An example that we can point to was on 16 February 2021, two weeks after the coup in Myanmar, the US Military's social media accounts were attacked by an inauthentic social media botnet. In this attack many social media accounts were inundated with comments containing text and images. On Facebook alone a single US Military Facebook page received 764,092 comments over a 24 hour period where the activity was occurring at 3,000 comments per minute. This attack was designed as an information operation to achieve information dominance on the topic over the given period. Of note Facebook and many suppliers marked the activity authentic and only through manual review using big data and mathematics could we assess that the entire activity was actually inauthentic.[1]

These inauthentic social media botnet attacks occur often and there is reliable and long reporting history of the use of these capabilities to achieve geopolitical objectives or influence elections by China and Russia.[2] Lynas Metals was advised by Mandiant (Google) that their accounts in April and May 2022 were attacked "in the Chinese national interest to incite

[1] https://www.afr.com/policy/foreign-affairs/inside-a-myanmar-coup-botnet-that-blasted-us-army-facebook-20220905-p5bfdw
[2] King, G., Pan, J., & Roberts, M. E. (2017). How the Chinese government fabricates social media posts for strategic distraction, not engaged argument. American Political Science Review, 111(3), 484-501.

protests…against the company's plan to build a rare earths processing plant in Texas"[3] During the Brexit referendum social bots accounted for a large proportion of the political debate on Twitter with the specific intent of spreading misinformation, contorting the true positions of parties and experts, and polarizing public opinion on the referendum.[4] During the 2016 US Presidential election Woolley and Guilbeault found that bots were employed to manipulate public opinion on both sides of the debate. [5]

The most sophisticated software that exists to create large inauthentic social media botnets in our assessment are developed by cyber criminals from Russia. This is because Russian cyber criminal organizations have been building Black SEO businesses to conduct advertising fraud and spam phishing campaigns for at least the past 20 years. In our assessment their capabilities have recently been adopted to enable foreign interference and vaccine disinformation through social media bot farms.[6] A good example of this is the SANA platform.

The SANA platform is an advanced social media botfarm software that we assess was the likely originator software for the Myanmar Coup Botnet attack on the US Miliatry in February 2021. The SANA interface serves as a gateway for a user to build and deploy inauthentic social media campaigns. Digital Revolution recently released a collection of twenty files, including a video and sixteen screenshots that walk users through the SANA interface[7]. The platform is capable of controlling the campaigns across 6 major social media sites and dozens of blogs. It is designed to deploy content in large newsbreaks amplifying messages or responding with positive, negative or neutral sentiment through comments and discussion. A behaviour model allows an operator to specify the times and frequency for bot activities, such as likes, comments, reactions, and group interactions. Response models dictate how a group of bots should react to news, and dictionaries categorize phrases, quotes, and comments as positive, negative, or neutral for use in social media responses. Finally, albums organize photograph sets for platform bot accounts. The system is built to bypass security measures as the operator selects names and surnames from a list and chooses an SMS API platform to create a phone number that can automatically respond to text requests.

In our assessment the two significant impacts to inauthentic social media campaigns by hostile foreign states are more division within society and a loss of trust in the results of elections. During the 2016 US Presidential election Russia conducted divisive campaigns on Facebook.[8] The Washington Post reported that Russia spent approximately USD$100,000 over 470 fraudulent Facebook pages conducting divisive messaging to both political sides focusing on

---

[3] https://www.afr.com/policy/foreign-affairs/miners-targeted-in-pro-china-cyberwar-claim-20220628-p5ax8p

[4] Howard, P. N., & Kollanyi, B. (2016). Bots, #Strongerin, and #Brexit: Computational propaganda during the UK-EU referendum. arXiv preprint arXiv:1606.06356.

[5] Woolley, S. C., & Guilbeault, D. R. (2017). Computational propaganda in the United States of America: Manufacturing consensus online. Computational Propaganda Research Project.

6 Krebs, B (2014) Spam Nation. https://www.amazon.com/Spam-Nation-Organized-Cybercrime_from-Epidemic/dp/1501210424

7 https://www.youtube.com/watch?v=bATLEMoi_h0
https://6068438 fs1 hubspotusercontent-na1 net/hubfs/6068438/fronton-report.pdf

[8] https://www.washingtonpost.com/business/technology/russian-operatives-used-facebook-ads-to-exploit-divisions-over-black-political-activism-and-muslims/2017/09/25/4a011242-a21b-11e7-ade1-76d061d56efa_story.html

pre-existing racial tensions. Dennis Yu, chief technology officer for BlitzMetrics, noted in the report that $100,000 worth of Facebook ads could have been viewed hundreds of millions of times. Putin has a long reporting history of using these methods and specifically focusing on preexisting racial tensions. In 2014 Putin told CBS's commenting on an incident involving the fatal shooting of Michael Brown, an African American, by a white police officer.

"If everything was perfect, there wouldn't be the problem of Ferguson. There would be no abuse by the police. But our task is to see all these problems and respond properly" Vladimir Putin

Michael A. McFaul, a former U.S. Ambassador to Russia noted that it is the long-term Russian objective to encourage discord in American society as he sees that if American society is imperfect, then our democracy is not better than his authoritarian rule. Sen. Mark R. Warner, then vice chairman of the Senate Intelligence Committee noted Russia intent was "to sow chaos" and "In many cases, it was more about voter suppression rather than increasing turnout." We see this voter suppression remark forewarning the future possible loss of trust in the results of elections following 2016.

In November 2017, The Times published an article which discussed how fake Twitter accounts posted more than 45,000 messages about Brexit in 48 hours in an apparent coordinated attempt to "sow discord", according to research by data scientists at Swansea University and the University of California, Berkeley. Notably, the tweets were mostly in support of Brexit, an outcome that Russia would have regarded as destabilizing for the European Union. However, a number of tweets were also in favor of Remain, indicating that the Russian goal may have been to sow division rather than promote a specific outcome. The use of bots in this way is part of a broader pattern of Russian interference in democratic processes around the world, we see this growing trend will at some point impact Australia in a large event.

Strategic risk exists with only a defensive strategy

Current platforms used by social media platforms to identify these inauthentic social media campaigns are inaccurate. Social media companies have a structural time and cost disadvantage in defending against state-sponsored actors intent to influence public opinion on their platforms because the foreign actors are not bound by domestic laws and are not restricted by conventional private-sector budgeting approaches and competing commercial priorities. Technology breakthroughs in generative artificial intelligence and "deep fakes" are making it harder over time to accurately detect inauthentic content. This increases the cost and complexity of bot monitoring and further drains those limited resources inside social networks to fight against bot influence.

Social media companies also in our opinion have strong financial disincentives to minimize bot influence, given that bots drive up engagement and usage metrics which in turn drives up company valuation and share prices. Furthermore, a large drop in usage and advertising metrics that follows a successful purge of bot accounts may be seen as an unpalatable risk that may cause a 'chain reaction' of advertisers and regular users to reconsider their commitment to the platform. As a result, foreign bot influence is not being adequately contained in shaping public opinion by social media companies.

There is no known regulation that requires social networks to act on bot activity in any specific timeframe or specific manner. We view it as unfair to the public that a bot can have the perceived equal view leading to a vote during a public debate. Voters get only one vote but group think on social media can influence voters to not participate if they perceive a large number of voters are talking against them in a hostile and non-inclusive manner. This is how voter suppression or disenchantment could impact Australia.

Suggested pillars to defend elections

Due to the significant cost associated with social media monitoring and accurately identifying bot activity, harnessing AI is probably the only way to conduct internet-wide surveillance and defend election integrity. We suggest the Government invests in AI-based identification platforms to identify, assess and flag attempts by inauthentic social media accounts to conduct foreign interference against voters.

We suggest the Government review its laws and regulations to enable the real-time identification, tagging and/or removal, in a fair but fast way, of inauthentic social media accounts influencing Australian elections and referendums. This can be augmented by Australian Government capabilities to go after the actors behind the software that conducts these campaigns. Social media companies, in our assessment, would need to be forced to cooperate with any measure to reduce the impact of active hostile campaigns.

Traditional media organisations should be carefully briefed and trained by the Australian Government or the private sector to ensure that they are not inadvertently reporting on or amplifying foreign influence on social media.

Awareness campaigns can have an inoculation effect when done well. The public currently understands bots but does not know how to identify them and can at times be misled by inauthentic accounts attempting to be real. Certain demographics in Australia have varying levels of social media literacy and this can be improved with awareness campaigns. Building awareness and community reporting campaigns can augment detection and assessment platforms the Government employs.

In our opinion any strategy to identify, assess and remove inauthentic social media accounts discussing elections must be done in a fair, transparent and fast manner. No system is perfect and the ability for fair recourse will help limit adverse reactions to accidental temporary suspension when real voters are using anonymous accounts to participate in public debates.