BALSILLIE SCHOOL
OF INTERNATIONAL AFFAIRS

Discord and Disruption

2019 Global Trends Report

An Anthology of Briefing Notes by Graduate Fellows at the Balsillie School of International Affairs

The Foreign Policy Research and Foresight Division at Global Affairs Canada is proud to support and be associated with the Graduate Fellowship Program/Young Thinkers on Global Trends Initiative. The challenges facing Canada today are unprecedented and truly global. Tackling those challenges require fresh ideas and engagement with new generations of thinkers, researchers, and activists to help create opportunities for a sustainable future. We would like to thank the students and professors of the Balsillie School of International Affairs for their time, effort and commitment throughout the year to make this initiative successful. The results of their work, which has been encapsulated in this anthology, will help inform the work of Global Affairs Canada as it relates to foreign policy, trade and international development.

Global Affairs Canada | Affaires mondiales Canada

BALSILLIE SCHOOL OF INTERNATIONAL AFFAIRS

67 Erb Street West
Waterloo, ON N2L 6C2 Canada
Telephone: 226 772 3001

The Role of Artificial Intelligence in Countering Online Violent Extremism

Rebecca Herbener, Sebastian Lacey and Matthew Markudis

Issue

This background note presents options to address the current and potential future threat posed to the rights and safety of the Canadian public by online extremist content and radicalization of vulnerable individuals.

Background

Challenges

The Internet allows domestic, foreign and state-sponsored extremists to widen their reach to Canadian citizens. However, online spaces are only one contributing factor to a person's decision to turn to violent extremism. Government has a key role to play in promoting national sovereignty and protecting the rights of citizens. In a globalized world where information transcends borders, the Internet is often exploited by radical groups to support their agendas, such as facilitating recruitment efforts, sustaining financial flows, logistics planning and contributing to the arms trade.

In the past few years, dominant technology firms such as Facebook, Twitter and YouTube have been used to spread propaganda and connect moderate users to extremists (Stevens and Neumann 2009). Following a series of terrorist attacks worldwide, this problem caught the attention of governments, who have put pressure on online platforms to address this issue. In response, these organizations have expanded and increased the sophistication of their artificial intelligence (AI) programs to identify and remove extremist content faster.

What Is AI?

AI is an area in computer science that emphasizes the creation of intelligent machines that work and react like humans (Grosz et al. 2016). It demonstrates many behaviours associated with human intelligence, including planning, learning, reasoning, problem solving, knowledge representation and social intelligence. Currently, most applications of AI are algorithms that are trained on datasets to recognize patterns and automate decision-making processes.

Increasingly, social media corporations are using AI to fight online extremism. While the effectiveness of these AI programs continues to increase and demonstrate their value, there have been many examples of false identification. For example, in the United States, face recognition programs have disproportionately flagged visible minorities as high-risk offenders in federal law enforcement systems (Goode 2018). These false positives are problematic, as they have the capacity to infringe on rights and hinder free speech. One high-profile instance of this was following Germany's adoption of a controversial new hate speech law in early 2018, which resulted in many cases of false positives, including a German satirical magazine's Twitter account being blocked after it parodied anti-Muslim comments (Bennett and Livingston 2018). These examples highlight the delicate balance that must be maintained between citizens' rights to free speech and national security.

While AI solutions to counter online extremism can be effective, lacking sufficiently comprehensive datasets can result in misidentification and bias. Despite progress, the

value added by AI remains in their capacity to automate repetitive tasks, rather than replace human analyst's expertise. As such, AI remains most effective when used in conjunction with humans to more accurately sort through the information and prevent false positives (Scott, Heumann and Lorenz 2017).

Legislation

Aggressive domestic legislation complements existing risk-adverse behaviour of dominant technology firms, which has led to over-policing of online platforms. These actions can have downstream consequences. An example of this is the flight of users away from major online platforms towards other online spaces that are unable or unwilling to adhere to commitments of corporate social responsibilities (Lomas 2017). Any future policy must be proactive by focusing on collaborating with the private sector instead of on regulations that restrict freedom of expression.

Private Sector Collaboration

The government could also support small and medium sized enterprises (SMEs) that often lack sufficient resources to fight extremist content on their platforms. Due to the increased allocation of resources from major tech companies towards countering violent extremism online, it is likely radical groups will turn to these smaller platforms and exploit their deficiencies. The United Nations Counter-Terrorism Committee Executive Directorate, in collaboration with the private sector, has created an initiative to support the tech industry to fight terrorist exploitation of their technologies called Tech Against Terrorism (Cohen 2017). Underneath its umbrella, it has launched the Knowledge Sharing Platform, which is a collection of tools that digital SMEs can use to better protect themselves. While this initiative is a step in the right direction, governments should further invest in these capacity-building initiatives to support SMEs so that they can defend their platforms against online extremism.

Unintended Consequences

There are three main potential consequences of using AI to counter online violent extremism. First, targeting specific groups such as Muslims or members of the alt-right, can unintentionally drive non-threatening members to extremism (Choudhury and Fenwick 2011). By creating a sense of surveillance and distrust of a specific community, it can foster suspicion leading to a perceived threat against that community. Second, removing specific content or specific users can push users to more encrypted sites of the Internet (Wilson 2016). This makes it more difficult to monitor and information becomes less accessible to intelligence and law enforcement agencies. Third, removing dissenting opinions and views can lead to a lack of non-violent avenues for expressing political and social grievances. It can also create echo chambers online that lead dissenting opinions to become extreme (Khan 2018).

The private sector needs to be encouraged to discriminate between bots that engage in actions to spread and amplify extremist content and those that merely automate tasks that humans are capable of completing. Shifting the focus from the nature of the disseminator, and instead towards their actions ensures that users retain the capacity to maximize their freedom of expression online while simultaneously protecting Canadian's Charter rights and freedoms. As the use of bots becomes more widespread, various groups such as academics and artists will continue to experiment with the legitimate use of this technology. There have been attempts by researchers to use bots to gather large quantities of data using keywords from social media platforms such as Twitter and Reddit. However, these attempts have been rightfully criticized as having flawed research methodologies, as current technologies are unable to discern whether content is genuine, sarcastic, or indicative of trolling behaviour. Furthermore, artists have used bots in order to create computer-generated poetry. While these attempts at using bots legitimately are largely imperfect, users will likely continue to innovate and find novel applications for this technology.

Next Steps

One role the government can take is that of a funding body, not to social media platforms, but to private AI firms that engage in data scraping, which allows for the mass collection of text-based characters. This would entail government collaboration with private enterprise, which would include government financing through staged payments to the company as they progress with the project. By funding private AI companies, the government helps encourage innovation without stifling ingenuity. As well, as an active financier, the government maintains oversight in the company's handling of Canadians' data and privacy. Finally, the government as a financer furthers current policy to make Canada a leader in AI technology.

Contracts between government and private companies can pursue AI in three main ways. First, algorithms can be programmed to begin compiling big datasets on online extremism since there is a lack of big data in this area. With this, AI and datasets will simultaneously grow, allowing for more advanced applications in the future. Second, algorithms can be reverse engineered. Instead of using previous online activity to flag potential future extremists, this method would trace the process through which individuals become radicalized. That is, instead of using an algorithm to predict the likelihood an individual will act on violent extremism, the algorithm will instead trace the steps a known violent extremism took to their point of radicalization. Finally, AI can be programmed to collect data in a two-step process. The first step would be broad information gathering. Then, on an ongoing basis, experienced analysts sort through the noise and identify immediate and emerging threats. At this point AI can then be programmed based on more narrow parameters to gain more information.

While a subsidy regime has many benefits, there exist several limitations. First, it has the capacity to incentivize digital SMEs, but its effectiveness with respect to the dominant technology firms with large operating budgets dedicated to AI development is questionable. Second, there is a potential for negative media coverage as Canada may be seen as picking winners and losers by providing capital to certain actors within this sphere. However, the Government of Canada in respect to contracts or procurement, routinely discriminates between firms. Ultimately, this is a challenge of communicating a future policy clearly to the Canadian public. Third, without sufficient public buy-in the policy may have a poor social reception, as there may exist a perception that Canada is engaging in corporate welfare to large technology firms for the development of technologies that remains suspect. Various government departments, including but not limited to the Department of National Defence, are recognizing the importance of AI. If Canada is to shape the future of AI, it needs to be an active stakeholder as opposed to a passive observer.

Recognizing that this issue is transnational, it would be beneficial to establish an international standard on online extremism among Canadian allies. Lacking such a standard, individual governments addressing this issue through different policies and regulations increases the operating cost required by companies to comply. An example of an attempt at creating an integrated digital regulation union is the European Union's General Data Protection Regulation (GDPR). Thus, negotiating a simplified set of standards would foster progress and efficiency, and provide the opportunity for governments to collaborate and share best practices. Despite these benefits, there are concerns that requiring digital companies to adhere to the GDPR reinforces norms of data protection and security that are unique to the European context.

Recommendations

- Government must be proactive and collaborate with private sector companies to invest in technology to counter violent extremism.
- The tools created with government support must be fully transparent in providing the public information on the purpose, scope, and results of these tools.
- Government should enact policy that protects citizens' data and digital rights while allowing private enterprise to innovate and flourish.

About the Authors

Rebecca Herbener is a student in the University of Waterloo's Masters in Global Governance program based at the BSIA.

Sebastian Lacey is a student in the University of Waterloo's Masters in Global Governance program based at the BSIA.

Matthew Markudis is a student in Wilfrid Laurier University's Master of International Public Policy program based at the BSIA.

Acknowledgements

The authors would like to thank Alistair Edgar for his guidance and mentorship as a supervisor throughout the development of this policy brief. Special thanks to the BSIA and Global Affairs Canada for their knowledgeable feedback and support throughout the course of this project.

Works Cited

Bennett, Lance and Steven Livingston. "The disinformation order: Disruptive communication and the decline of democratic institutions." *European Journal of Communication* 33 (2): 122–39. doi: 10.1177/0267323118760317.

Choudhury, Tufyal and Helen Fenwick. 2011. "The Impact of Counter-Terrorism Measures on Muslim Communities." *International Review of Law, Computers, and Technology* 25 (23): 151-181. doi: 10.1080/13600869.2011.617491.

Cohen, David. 2017. "The global internet forum to counter terrorism gets going today." *Adweek*, August 1. www.adweek.com/digital/global-internet-forum-to-counter-terrorism-workshop-san-francisco/.

Goode. 2018. Facial Recognition Software is Biased Towards White Men, Researcher finds." *The Verge.* https://www.theverge.com/2018/2/11/17001218/facial-recognition-software-accuracy-technology-mit-white-men-black-women-error

Grosz, Barbara, Russ Altman, Eric Horvitz, and Alan Mackworth. 2016. "Artificial Intelligence and Life in 2030." *Stanford University Report on the 2015 Study Panel.*

Khan, Adnan R. 2018. "How the Internet may be Turning us all into Radicals." *Macleans,* June 26. www.macleans.ca/society/technology/how-the-internet-may-be-turning-us-all-into-radicals/.

Lomas, Natasha. 2017. "Germany's social media hate speech law is now in effect." *Tech Crunch*, October 2. https://techcrunch.com/2017/10/02/germanys-social-media-hate-speech-law-is-now-in-effect/.

Scott, Ben, Stefan Heumann, and Philippe Lorenz. 2018. "Artificial Intelligence and Foreign Policy." *Stiftung Neue Verantwortung*.www.stiftung-nv.de/de/publikation/artificial-intelligence-and-foreign-policy.

Stevens, Tim and Peter Neumann. 2009. "Countering Online Radicalisation: A Strategy for Action." *The International Centre for the Study of Radicalization and Political Violence.*

Wilson, Jason. 2016. "Gab: Alt-Right's Social Media Alternative Attracts Users Banned From Twitter." *The Guardian*, November 17. www.theguardian.com/media/2016/nov/17/gab-alt-right-social-media-twitter.