Justice and Human Rights Committee
Online Hate Study
Brief Submission

Combatting Online Hate: An Alternative Approach

Alyssa Blank, M.A.

Alyssa Blank has a Master of Arts degree in Globalization and International Development from the University of Ottawa, where she focused her studies on online hate and the psychological operations of war. She has an extensive background in analyzing and addressing online and offline hate.

Executive Summary

As a means of effectively combatting online hate, this brief brings together online and offline actions, working with the data that exists through social media companies and within their existing terms and conditions to propose an alternative to removing offensive materials from online sources. It proposes that the Canadian government should aim to:

- Work with social media companies to identify the geographic locations of the online offensive materials, as flagged through user reporting;
- Work with experts to identify social realities and/or tensions in these areas offline; and
- Implement tailored programming for these locations.

This structure will create a real-time metric of evaluation for countering violent extremism (CVE) programming—something practitioners currently lack—while encouraging social media companies' development of practical innovation for the content they are currently collecting and sharing.

Grappling with Un-checked Online Hate

The complex discussion around how to combat online hate has been brought into sharp focus in recent months, as the world bore witness to attacks on religious communities, beginning with the targeting of Jews at prayer in the United States in October 2018. Just as faith communities were catching their breaths, Muslims at prayer were targeted in New Zealand, Christians at prayer were targeted in Sri Lanka and Jews at prayer, once again, were targeted in the United States. Most of these attacks carried an online component which left many trying to figure out how to moderate online content to safeguard society, while protecting freedom of expression.

Many of these concerns have been directed towards social media companies, the oft-unwitting hosts to these violent or hateful postings. With international efforts to try and take down these posts, or prevent their presence to begin with, the underlying narrative is that the internet has become unruly and is fomenting hatred. Although it can be argued that the internet and its social media companies allow a level of unchecked freedom to hate, it is equally true that the internet is a reflection of its users' real-time impressions and experiences in the world. Indeed, it is essential to recognize that offline hate finds expression online, which can be an impetus for, or part of, a plan for real-world violence. In this way, online hate does not exist in a vacuum and, instead, is a reflection of external social factors, which can vary geographically.

While hateful expressions online reasonably result in a desire for removal, they can instead be seen as an opportunity to work with social media platforms on which these posts are shared, to explore, identify and effectively address the reality of dangerous hate where it truly lives: offline. With this in mind, this brief proposes that to combat online hate effectively, the Canadian government has the opportunity to enact a multi-pronged approach in collaboration

with social media companies to learn from each other and create a way to bridge online and offline realities to address online and offline hate.

European Models and Challenges

Multiple initiatives to combat online hate exist or are being explored throughout the world, with European models providing considerable lessons for Canada. The broad and opt-in approach of the European Union, the strict and far-reaching approach in Germany and the burgeoning holistic proposal currently being investigated in the United Kingdom.

In 2016, the European Commissioner established a *Code of Conduct on Countering Illegal Hate Speech Online*, which requires willing online companies to assess user reports flagging offensive materials within 24 hours and then, in following with EU and national legislation, take down illegal content[1]. While covered by the EU's approach, Germany enacted its Network Enforcement Act (NetzDG) in 2018. NetzDG requires social media companies with more than 2 million German users to take down offensive materials online, in accordance with German law, in 24 hours (or, for more complicated cases, within 7 days) or face fines of up to 50 million euros[2]. Finally, as of April 2019, the UK is seeking a multi-pronged regulatory framework to ensure online safety through an ongoing consultative process. This approach broadly seeks to establish oversight and collaborate with social media companies.

The general critique of approaches which emphasize the removal of online material is that doing so is a violation of freedom of expression and that these restrictions to a fundamental democratic right is at the discretion of privately-owned social media companies. This is a concern that is held by the public and social media companies alike. Indeed, in regards to NetzDG, Facebook's Vice President of Communications and Public Policy, Elliot Schrage, stated that "…'we think it's a bad idea for the German government to outsource the decision of what is lawful and what is not'" [5].

With this in mind, Canada has the opportunity to learn from peer nations that have already implemented models specific to their unique jurisdictions, as well as internalize the lessons from the challenges that have arisen to develop a made-in-Canada approach.

A Way Forward: A Canadian Collaborative Approach

Foundational to any effective approach to CVE is the recognition that online hate is a reflection of offline reality. As such, removing hateful posts as a central means of addressing hate will create a never-ending endeavor that fails to address the root cause of the issue. Furthermore, punishing social media companies for being adversaries of social well-being—through fines or

[1] European Commission. 2019. Countering illegal hate speech online – EU Code of Conduct ensures swift response. European Commission. http://europa.eu/rapid/press-release_IP-19-805_en.htm

[2] Dodds, Laurence. 2018. British MPs call for German-style law to block hate speech on social media. The Telegraph. https://www.telegraph.co.uk/technology/2018/07/28/british-mps-call-german-style-law-block-hate-speech-social-media/

legislation—is a short-sighted approach that misses the opportunity to address hate speech effectively.

An effective approach to address online and offline hate, therefore, should be based on a collaboration between the government of Canada and social media companies to:

- Identify hate speech;
    - The government of Canada must define what this means.
    - Social media companies must assist the government in tracking this, with governmental oversight.
- Identify the real-world location of hate speech online;
    - The motivation behind hate speech can vary from region to region. Identifying where hate speech is from, will allow insight into why this is happening and how it can be addressed.
    - Social media companies collect and share similar information currently, while maintaining the anonymity of users (see Appendix 1).
- Analyze the regional realities of these posts;
    - CVE practitioners can investigate the social factors in the regions identified by location information collected by social media firms.
- Create programming to address key factors identified by CVE practitioners, thereby addressing online hate; and
- CVE practitioners can assess the effectiveness of the programs with the assistance of data provided by social media companies.

Hateful posts on social media and the reporting of them highlight tension—about which social media companies have collected data and location information—that can be shared with governments much in the same way social media sites share content preferences with advertisers (see Appendix 1).

Collecting this data would allow the government to build a robust understanding of the regional realities from which these posts emanate. Factors such as: unemployment rates, resource allocation, poverty levels, crime rates, multiculturalism, education, social isolation, etc. can then be analyzed offline and tailored programs can be put into place to curtail the increase of online hate and ultimately offline action.

Addressing hate in this way will allow CVE practitioners to implement a metric of evaluation through the establishment of a baseline of tension, as identified through the data provided by social media companies, all the while maintaining users' privacy and respecting freedom of expression.

Appendix 1: Relevant Sections of Social Media's Terms of Service and Privacy Policies

Currently, Facebook, Twitter and Instagram have a default setting, of which users can opt out, to collect geographic data which is then shared with certain bodies, including advertisers. These social media outlets also have a Terms of Service component, which allows for that sharing of information with governments, as needed, and outline their approach to sharing information for security purposes. The applicable passages are as follows:

Facebook and Instagram

- According to Facebook's Privacy Policy, they "…use location-related information – such as your current location, where you live, the places you like to go, and the businesses and people you're near – to provide, personalise and improve our Products, including ads, for you and others. Location-related information can be based on things such as precise device location (if you've allowed us to collect it), IP addresses and information from your and others' use of Facebook Products (such as check-ins or events you attend)"[3].
- Facebook notes that this information is used to enhance advertising, promote safety and conduct research and innovation for the purpose of improving the social good. They specify:
    - "We use the information that we have to verify accounts and activity, combat harmful conduct, detect and prevent spam and other bad experiences, maintain the integrity of our Products, and promote safety and security on and off Facebook Products. For example, we use data that we have to investigate suspicious activity or breaches of our Terms or Policies, or to detect when someone needs help."
    - "We use the information that we have (including from research partners who we collaborate with) to conduct and support research and innovation on topics of general social welfare, technological advancement, public interest, health and well-being. For example, crises to aid relief efforts."

Twitter

- According to its Terms of Service, Twitter reserves "…the right to access, read, preserve, and disclose any information as we reasonably believe is necessary 10 (i) satisfy any applicable law, regulation, legal process or governmental request, (ii) enforce the Terms, including investigation of potential violations hereof, (iii) detect, prevent, or otherwise address fraud, security or technical issues, (iv) respond to user support requests, or (v) protect the rights, property or safety of Twitter, its users and the public. Twitter does not disclose personally-identifying information to third parties except in accordance with our Privacy Policy[4].
- According to its Privacy Policy, Twitter collects location data, based on the user's privacy settings, and can share this information with any number of interested parties. It notes

[3] Facebook Privacy Policy https://www.facebook.com/about/privacy/update; Instagram Privacy and Safety Centre https://help.instagram.com/581066165581870

[4] Twitter Terms of Service https://twitter.com/en/tos

that “Subject to your settings, we may collect, use, and store additional information about your location - such as your current precise position or places where you’ve previously used Twitter - to operate or personalize our services including with more relevant content like local trends, stories, ads, and suggestions for people to follow.”[5]

- The privacy policy goes on to outline when they will share private information by stating: “Notwithstanding anything to the contrary in this Privacy Policy or controls we may otherwise offer to you, we may preserve, use, or disclose your personal data if we believe that it is reasonably necessary to comply with a law, regulation, legal process, or governmental request; to protect the safety of any person; to protect the safety or integrity of our platform, including to help prevent spam, abuse, or malicious actors on our services, or to explain why we have removedcontent or accounts from our services; to address fraud, security, or technical issues; or to protect our rights or property or the rights or property of those who use our services. However, nothing in this Privacy Policy is intended to limit any legal defenses or objections that you may have to a third party’s, including a government’s, request to disclose your personal data.”

[5] Twitter Privacy Policy https://twitter.com/en/privacy