

Report Highlights: Processes, People, and Public Accountability

Centre for Media, Technology and Democracy

A new report by Chris Tenove and Heidi Tworek outlines the complexity of harmful online communication:

“it includes factors such as algorithms and interface designs, rapidly changing use patterns, and interactions between different digital services. These new factors complicate the longstanding challenge of determining how messages and media affect people’s beliefs and behaviors. It is difficult to conceptualize and measure harms from communication, but doing so is crucial for mitigating those harms and balancing any interventions against restrictions to free expression and other goods.”

Report Highlights

During the 2019 Canadian federal election, three forms of harmful online communication were observed: abuse of individuals; intolerance and hate toward marginalized groups in public online spaces; and building support for hate in private online spaces. The forms of online communication included types of speech that are already illegal in Canada, some that are harmful but not illegal, and harmful patterns that contribute to systemic discrimination. - Identifying online hate speech can be difficult for human evaluators and algorithms. The language itself is dynamic and contextual, employing new terms and euphemisms; combining insults with positive messages; and using the same term with a different meaning in different contexts. The use of algorithms in identifying online hate speech can be particularly difficult because they use data labelled by humans and can inherit their biases or limitations. Algorithms can also miss important contextual information in conversations when looking at singular messages. - Since actors promoting discrimination and hate use multiple platforms, policies limited to a single platform will likely be ineffective. Policymakers need to consider the relationship between online and offline messages. - The study analyzed approximately one million tweets directed at candidates in the 2019 Canadian federal election between mid-August and October 31, and interviewed over 30 of the candidates or their communication staff. They broke the negative messages into three categories: - Low negativity, which are dismissive or disrespectful. - Medium negativity, which are insulting or advance negative stereotypes. - High negativity, which include “hateful language at social groups, threats, unsubstantiated accusations of moral or criminal wrongdoing.” - Initial findings suggest over 40% of tweets fall into one of those three categories. 27% of the tweets in the sample were identified as low negativity, 13% as medium negativity, and 1% as high negativity. - Candidates report that high negativity messages “affected their sense of wellbeing and security for themselves and their staff, and could require time-consuming engagements with police or civil servants,” while low and medium negativity messages could be demoralizing for them and their staff. - Campaign teams developed ad hoc and reactive strategies to manage the abuse they were receiving, while candidates demanded greater accountability for the people who post these messages and expressed frustration at the limited actions that they could take in response to them. They also believed that social media companies should face greater accountability for the abuse and hate they allow on their platforms. - The authors recommend a policy response that focuses on processes, people, and accountability, instead of on individual pieces of content. They suggest the creation of a social media council to develop principles to address harmful communication in the Canadian context; greater support for candidates to manage the threats

they receive; efforts to understand and educate those who promote hate in public and private spaces; and greater access to independent research on data from social media platforms.

While this study examines how hate online communication is specifically deployed toward electoral candidates, it's experienced by a far greater number of online users and has clearly become something that needs to be addressed to ensure our online spaces can be civil and inclusive. That will not be accomplished simply by waiting for platforms to act; rather, it will require a policy response. The authors give examples of the form such a response has taken elsewhere in the world, and begin to lay out a framework for what an effective policy response might look like in Canada.