Data&Society

**This case study was utilized at an AI and Human Rights workshop, held at the Data & Society Research Institute on April 26-27, 2018.*

Social Media, Artificial Intelligence, and Hate Speech in Myanmar
Case Study

Background

In March 2018, the UN Independent International Fact-Finding Mission on Myanmar reported "concrete and overwhelming"[1] evidence of human rights violations against the Rohingya population. United Nations investigators concluded that government and military efforts to persecute the community in all likelihood amount to crimes under international law.[2] Investigators found that social media, in particular Facebook, has played a "determining role"[3] in spreading hate speech and disinformation, which further incited violence against the ethnic minority.

On April 10, 2018, during congressional testimony, Senator Patrick Leahy questioned Mark Zuckerberg on the issue and asked why these postings couldn't be removed within 24 hours. He cited the repeated requests that Myanmar civil society groups made to Facebook's team to stop their dissemination.[4] Zuckerberg responded by pledging to do more to combat dangerous speech, including hiring dozens more Burmese-language content reviewers and making product changes in the country. He also continuously highlighted Facebook's use of artificial intelligence tools to combat dangerous content.[5]

While some civil society groups welcomed Zuckerberg's commitment to tackle hate speech,[6] they explained how this case "reveals an over-reliance on third parties, a lack of a proper mechanism for emergency escalation, a reticence to engage local stakeholders around systemic solutions and a lack of transparency."[7] They contended that a reliance on AI tools as a main measure to detect and respond to this issue is not an adequate response and that more resources are still urgently needed. As quoted in the Washington Post, Robyn Caplan of Data & Society criticized Facebook's claims that AI can solve such problems, saying "AI can't understand the context of speech and, since most categories for problematic speech are poorly defined [by necessity], having humans determine context is not only necessary but desirable."[8]

---

[1] "Fact-finding Mission on Myanmar: concrete and overwhelming information points to international crimes." March 12, 2018. http://www.ohchr.org/EN/HRBodies/HRC/Pages/NewsDetail.aspx?NewsID=22794&LangID=E

[2] "Statement by Mr. Marzuki Darusman, Chairperson of the Independent International Fact-Finding Mission on Myanmar, at the 37th session of the Human Rights Council." March 12, 2018. http://www.ohchr.org/EN/HRBodies/HRC/Pages/NewsDetail.aspx?NewsID=22798&LangID=E

[3] Reuters. "Myanmar: UN blames Facebook for spreading hatred of Rohingya." March 12, 2018. https://www.theguardian.com/technology/2018/mar/13/myanmar-un-blames-facebook-for-spreading-hatred-of-rohingya

[4] Bloomberg Government. "Transcript of Mark Zuckerberg's Senate hearing." April 10, 2018. https://www.washingtonpost.com/news/the-switch/wp/2018/04/10/transcript-of-mark-zuckerbergs-senate-hearing/?noredirect=on&utm_term=.2c1a3a175e57

[5] Ibid., note 4

[6] Andy Sullivan, Yimou Lee. "Myanmar activists welcome Zuckerberg's 24-hour target to block hate speech on Facebook." April 10, 2018. https://www.reuters.com/article/us-facebook-privacy-myanmar/myanmar-activists-welcome-zuckerbergs-24-hour-target-to-block-hate-speech-on-facebook-idUSKBN1HI028

[7] Libby Hogan. "Myanmar groups criticise Zuckerberg's response to hate speech on Facebook." April 5, 2018. https://www.theguardian.com/technology/2018/apr/06/myanmar-facebook-criticise-mark-zuckerberg-response-hate-speech-spread

[8] Drew Harwell. "AI will solve Facebook's most vexing problems, Mark Zuckerberg says. Just don't ask when or how." April 11, 2018. https://www.washingtonpost.com/news/the-switch/wp/2018/04/11/ai-will-solve-facebooks-most-vexing-problems-mark-zuckerberg-says-just-dont-ask-when-or-how/?utm_term=.85a6138cdb26

Ethical dilemmas

Facebook is the most popular social networking site in the world, with more than 1.8 billion active users per month.[9] In Myanmar it has "become a near-ubiquitous communications tool, following the opening up of the economy."[10] Given its far reaching impact, the platform's misuse to spread dangerous speech, has in effect, helped to perpetuate the institutionalized discrimination against the Rohingya community, who are often described as "the most persecuted minority in the world."[11] The challenges of identifying and removing anti-Rohingya propaganda demonstrates that an overreliance on AI tools cannot mitigate this problem alone. While AI can aid in flagging questionable content, it cannot be trusted to remove it.[12] The nuances of hate speech and linguistics may simply be beyond AI's current capabilities.[13] Even if AI does become more reliable in detecting and removing dangerous speech in the future, the algorithms, models, and data sources themselves will not be free from bias.

This case is important because it shows what's at stake for other communities if sustainable and rights-respecting solutions are not developed. For example, in Sri Lanka, the government recently blocked Facebook in an attempt to thwart violence against Muslim communities, and in Indonesia, politicians called on Facebook executives to account for the spread of disinformation.[14] The case further highlights the relationship between the private sector and UN mechanisms and the extent to which companies engage with the UN Guiding Principles on Business and Human Rights. It illustrates how companies may need to be more proactive and conscientious about assessing and remedying the potential human rights impacts of their products. By actively working to adhere to this UN framework, companies can indicate a willingness to fulfill their social and normative responsibilities to respect human rights.

Human rights implications

While freedom of expression implies the right to scrutinize, debate, and criticize ideas, opinions, and beliefs, even if harshly or unreasonably, it "does not advocate hatred that incites hostility, discrimination or violence against an individual or a group of individuals."[15] There are international standards that ban certain speech on the basis that it infringes on the equality and rights of others,[16] including freedom from discrimination and fear, freedom of religion and belief, and the right to self-determination.

However, as Lucy Perdon argues, many challenges arise with legislating against hate speech online "because it is not always easy to distinguish where freedom of expression ends and legitimate restriction on expression begins." Further, the speech in question "may be region- or culture-specific, rooted in a country's history. The lack of an internationally agreed definition of hate speech has made it difficult to clarify how such acts should be dealt with, including in the digital world." Therefore, a balanced approach

---

[9] Rosamond Hutt. "The world's most popular social networks, mapped." March 20, 2017. https://www.weforum.org/agenda/2017/03/most-popular-social-networks-mapped/

[10] Ibid., note 3

[11] Gabriella Canal. "Rohingya Muslims Are the Most Persecuted Minority in the World: Who Are They?" February 10, 2017. https://www.globalcitizen.org/en/content/recognizing-the-rohingya-and-their-horrifying-pers/

[12] Ibid., note 9

[13] Larry Greenemeier. "Can AI Really Solve Facebook's Problems?" April 13, 2018. https://www.scientificamerican.com/article/can-ai-really-solve-facebooks-problems1/

[14] Kevin Roose, Paul Mozur. "Zuckerberg Was Called Out Over Myanmar Violence. Here's His Apology." April 9, 2018. https://mobile.nytimes.com/2018/04/09/business/facebook-myanmar-zuckerberg.html

[15] Report of the Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression. "Promotion and protection of the right to freedom of opinion and expression." August 10, 2011. http://www.ohchr.org/Documents/Issues/Opinion/A.66.290.pdf

[16] Toby Mendel. "Hate Speech Rules Under International Law." February 2010. http://www.law-democracy.org/wp-content/uploads/2010/07/10.02.hate-speech.Macedonia-book.pdf

must be taken to ensure that any efforts to restrict speech "do not stifle free expression and are aligned with international human rights law."[17]

Discussion questions

1. Facebook has pledged to take specific actions to combat hate speech and disinformation in Myanmar. What steps can tech companies take to respond in similar cases and hold themselves accountable?
2. In Mark Zuckerberg's congressional testimony, elected officials questioned him regarding the social platform's role in fueling the Rohingya crisis. How can governments effectively call for corporate accountability and transparency?
3. In thinking about ethics by design, should ethics training be offered to AI engineers and developers? How would this be carried out? What are some current examples?
4. How have civil society and corporations engaged on the topic of hate speech and disinformation? What are some best practices or past examples? What are the biggest challenges?
5. Do the UN Guiding Principles on Business and Human Rights sufficiently outline the responsibility of corporate actors in a case such as the one described? Are new principles or protocols required to guide private actors in specifically assessing the human rights impact of artificial intelligence?

Case study prepared by Melanie Penagos

[17] Lucy Purdon. "The Challenge of Criminalising Hate Speech." August 16, 2016. https://www.ihrb.org/focus-areas/information-communication-technology/challenge-of-criminalising-hate-speech