Online Xenophobic Behavior Amid the COVID-19 Pandemic: A Commentary

ROY KA-WEI LEE, Singapore University of Technology and Design

ZHI LI, University of Saskatchewan

The COVID-19 pandemic has developed to be more than a bio-crisis as global news has reported a sharp rise in xenophobia and discrimination in online and offline communities. This commentary discusses the deterrence policies and automated content moderation technology to curb the spread of online xenophobic behaviors. We suggested some potential exploration areas where multiple stakeholders can come together to address the online xenophobic issues amid the COVID-19 pandemic.

CCS Concepts: • **Social and professional topics** → **Government technology policy**; • **Computing methodologies** → **Artificial intelligence**; • **Applied computing** → **Law, social and behavioral sciences**;

Additional Key Words and Phrases: Social media, xenophobic behavior, hate speech

ACM Reference format:

Roy Ka-Wei Lee and Zhi Li. 2020. Online Xenophobic Behavior Amid the COVID-19 Pandemic: A Commentary. *Digit. Gov. Res. Pract.* 2, 1, Article 2 (November 2020), 5 pages.

<https://doi.org/10.1145/3428091>

1 BACKGROUND

In early March 2020, a video went viral online, showing an Asian man desperately hailing a ride on a highway after an Uber driver dumped him on Cairo's ring road on suspicion of having the coronavirus [25]. A voice heard in the video jokingly shouts: "The first coronavirus case in Egypt!" The same voice then tells the driver: "May God support you, Hajji! Throw him out!" The driver reportedly forced the foreigner out of the car later on the often-busy highway in Cairo. The viral video sparked comments from both camps: outrage against the racism act and online xenophobic remarks against the Chinese community.

The coronavirus disease 2019 (COVID-19) pandemic, which is the ongoing pandemic of cCOVID-19 caused by the severe acute respiratory syndrome coronavirus 2 [18], has seen more than 11 million reported diagnosed cases and more than 500,000 deaths globally as of July 2020. Globally, countries have taken unprecedented measures such as social distancing, contact tracing, and border closures to curb the spread of the coronavirus. However, the COVID-19 pandemic has developed to be more than a bio-crisis as a worldwide sharp rise in xenophobia and discrimination in online and offline communities has been reported [1, 7]. Previous studies had shown that ethnic outgroups are often blamed for outbreaks of epidemic diseases, and these outbreaks can inspire violently

Authors' addresses: R. K.-W. Lee, Singapore University of Technology and Design; email: roy_lee@sutd.edu.sg; Z. Li, University of Saskatchewan; email: z.li@usask.ca.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

© 2020 Association for Computing Machinery.

2639-0175/2020/11-ART2 \$15.00

<https://doi.org/10.1145/3428091>

xenophobic reactions to outsiders [9, 16]. Specifically, amid the COVID-19 pandemic, the hostility and discrimination toward Chinese communities in online social media have increased as China was presumed to be the ground zero of the coronavirus [23]. A white paper [31] from the Network Contagion Research Institute, a New Jersey-based nonprofit group that tracks hate speech online, also documented a sharp increase in xenophobic messages against the Chinese community on multiple online social media platforms. The report mentioned an Instagram post that called for mass shooting of Asians in Chinatown to end the COVID-19 epidemic in New York City. Such appalling calls for mass violence against a community raise the alarm for immediate actions to be taken to address this pressing and growing online xenophobia issue.

The spread of online xenophobic behaviors is not unique to the COVID-19 pandemic. The analysis and detection of online xenophobic behaviors such as hate speeches [10, 24] has been widely studied by the social, computer, and data science research community. The spread of antisocial behaviors in social media has not only sown discord among individuals or communities online but also has resulted in violent hate crimes [17, 22, 30]. Therefore, it is a pressing issue to detect and curb online antisocial behaviors, particularly during a pandemic, where preventing such undesirable online behaviors can avoid adding problems to the already difficult crisis. Despite the gravity of the issue, governments and social media platform providers have taken inadequate measures to tackle online xenophobic behaviors. This commentary aims to outline the potential areas of improvement and call the stakeholders to take active actions to deter and moderate online xenophobic behaviors.

2 DETERRENCE TO ONLINE XENOPHOBIC BEHAVIORS

Hate speech presents complex policy challenges, as it often brings two fundamental rights into conflict: equality, including the right to be protected from discrimination, and freedom of expression [12]. As such, the world's democracies fiercely disagree on the law and policies in regulating hate speeches. Consider the *Criminal Code* of Canada: there are three separate hatred-related offenses [5]: section 318 (advocating genocide), section 319(1) (publicly inciting hatred likely to lead to a breach of the peace), and section 319(2) (willfully promoting hatred). Although the details differ, the legislation of this sort exists in the preponderance of developed democracies, including Australia, Denmark, France, Germany, India, South Africa, Sweden, and New Zealand [13]. Yet such legislation would be struck down as unconstitutional in the United States, as an attack on free speech. The proliferation of social media further compounded the policy challenges of addressing xenophobic behaviors. First, social media enable users to spread messages, including xenophobic content, across national borders. This opens up a floodgate of complexity dealing with international laws to moderate the hateful and discriminatory content. Second, social media platforms offer anonymity and have a responsibility to protect their user's data privacy. This adds challenges to the investigations of online hate crimes.

With this complex and challenging policy backdrop for tackling online xenophobic behaviors, it is unsurprising that countries have not taken many actions to deter the spread of online xenophobic behaviors amid COVID-19, even though the cases of xenophobic conducts were reported to have risen sharply [1, 7]. The United Nations (UN) stood up to the challenge of campaigning for deterrence to xenophobic behaviors and protecting human rights amid the COVID-19 pandemic [19, 20]. As xenophobia spreads online and people of Asian descent have been stigmatized for spreading the coronavirus, many have endured the verbal and physical assaults and, at times, institutional exclusion from the receiving society. It is also noted by the World Health Organization that stigma can drive people to hide their illness to avoid discrimination, preventing them from seeking immediate health care. To address this aspect of the crisis, the International Organization for Migration issued a public communication toolkit that promotes placing participatory approaches to encourage Asian immigrants to communicate their problems. Among the recommendations for a variety of entities including member states and social media companies, the UN Guidance asks the member states to conduct independent investigation of allegations of hate speech and hold perpetrators accountable. Nevertheless, it is noteworthy that concrete actions are rarely taken or reported at government levels. With regard to social media companies,

the UN Guidance recommends that they should “take down COVID-19 related hate speech when it amounts to incitement to hostility, discrimination or violence, in accordance with international human rights law” [21]. Nevertheless, it is not easy to evaluate member countries and social media companies’ receptiveness and the downstream impact of these initiatives and recommendations.

A collective effort is needed to improve the policies for handling online xenophobia. We advocate joint collaboration between governments, non-government organizations (NGOs), and social media companies to deter online xenophobic behavior amid the COVID-19 pandemic. New working committees can be set up with representatives from the different stakeholders coming together to discuss the policies and solutions to tackle this pressing issue. For example, NGOs and governments can partner with social media companies to create advertising campaigns to raise awareness on the online xenophobic issues and promote social cohesiveness. NGOs that have championed human rights can actively identify and report new trends of xenophobic behaviors to social media companies so that the social media companies can factor these new trends in their detection algorithms. We believe that it is possible to strike a balance between protection of free speech and ensuring a safe environment for online social platforms users.

3 MODERATION OF ONLINE XENOPHOBIC BEHAVIORS

Major social media platforms such as Facebook and Twitter have made great efforts to combat the spread of hate speech in their platforms [4, 26]. For instance, the platforms have provided clear policies on hateful conducts [8, 28], implemented mechanisms for users to report hate speech, and employed content moderators to detect hate speeches actively. However, such approaches are labor intensive, time consuming, and thus not scalable or sustainable in the long run [11, 29]. To address this problem, many research studies have proposed traditional machine learning methods and deep learning models to detect and moderate online xenophobic behaviors and hateful content automatically [10, 24]. Most of these existing methods adopted a supervised approach where features are extracted from the textual content to train a classifier to predict if a given text contains hate speeches or xenophobic behaviors. These supervised methods have achieved promising results in detecting xenophobic content automatically. However, a large amount of annotated data is required to train the classifiers, and such annotated datasets are scarce. Furthermore, annotators may introduce biases in the annotation process and the biases are propagated in the classifier training, resulting in inaccurate xenophobic behaviors detection [2, 3].

The limitations of supervised xenophobic content detection methods become more salient when applied to examine online xenophobic behaviors amid the COVID-19 pandemic. Schild et al. [23] performed an empirical analysis of the emergence of sinophobic behavior on the web during the COVID-19 pandemic outbreak. The researchers focused their analysis on data collected from Twitter and 4Chan and found a new discriminatory lexicon specific to the COVID-19 pandemic. For example, *Kungflu* and *ChineseVirus* are new keywords that discriminate against the Chinese community. The new hate speech, discriminatory, and xenophobic words and phrases lexicon that evolved during the pandemic adds challenges to applying existing supervised machine learning methods for online xenophobic content detection. In particular, more annotations will need to be done to include the new xenophobic lexicon to train the classifier. Ziems et al. [32] attempted to fill this research gap by annotating the anti-Asian hate and counter-hate behaviors in a large dataset that comprises 30 million COVID-19 tweets. Nevertheless, the annotation is a laborious process, and there is a need to patch the xenophobic lexicon consistently.

The existing studies on online xenophobic behaviors during COVID-19 highlighted the room for improvements to develop more effective automatic moderation systems. Broadly, we suggest two main areas of improvement: (1) the exploration of unsupervised or semi-supervised xenophobic detection methods [15], and (2) the recommendation of xenophobic corrective and education material. One of the main insights gathered from the study of Schild et al. [23] is that we need to move away from a supervised approach in detecting online xenophobic content. A potential solution is to design unsupervised or semi-supervised machine learning models that

can perform xenophobic content detection by drawing inferences from the dataset without relying on the text's labels. For example, label-independent linguistic cues such as sentiment and topics can be used to auxiliary information for xenophobic content detection [6], or a clustering approach can be applied to find communities that are more susceptible to xenophobic content. Another direction is to consider the recommendation of xenophobic corrective and education material. Most of the existing studies have focused on detecting the hate speech and xenophobic content in the text while neglecting the post-moderation follow-up to educate and correct their xenophobic behaviors. To fill this research gap, persuasive human rights education content can be recommended to correct users who have exhibited online xenophobic behaviors [14].

4 SUMMARY: CALL TO ACTION

On March 11, 2020, the Uber driver who forced the Chinese national out of the car was arrested [27]. A follow-up video was also posted showing a group of Egyptians visiting the bullying victim in his hotel, offering him flowers and apologizing in Arabic and Chinese. The widespread condemnation of the racism act by social media users and the kind gesture of the Egyptians show that most social media users are sensitive and against xenophobic behaviors. Therefore, more could be done to maintain a safe online environment for social media users.

Deterring and moderating online xenophobic behaviors amid the COVID-19 pandemic is a complex and challenging task. Yet the COVID-19 pandemic also offers us an opportunity to develop the solutions to handle online xenophobic behaviors in future crises. To seize this opportunity, countries, NGOs, academia, and social platform providers have to work together to design and develop the policies and technologies need to combat the spread of online xenophobic behaviors amid the COVID-19 pandemic. It is our hope that this commentary provides some ideas and potential areas of exploration for the various stakeholders.

REFERENCES

- [1] UN News. 2020. COVID-19 stoking xenophobia, hate and exclusion, minority rights expert warns. *UN News*. Retrieved October 19, 2020 from <https://news.un.org/en/story/2020/03/1060602>.
- [2] Md. Rabiul Awal, Rui Cao, Roy Ka-Wei Lee, and Sandra Mitrović. 2020. On analyzing annotation consistency in online abusive behavior datasets. arXiv:2006.13507
- [3] Pinkesh Badjatiya, Manish Gupta, and Vasudeva Varma. 2019. Stereotypical bias removal for hate speech detection task using knowledge-based generalizations. In *Proceedings of the World Wide Web Conference*. 49–59.
- [4] Bloomberg. 2019. Twitter, Facebook Join Global Pledge to Fight Hate Speech Online. Retrieved February 11, 2020 from <https://www.bloomberg.com/news/articles/2019-05-15/macron-ardern-to-meet-twitter-facebook-google-on-hate-speech>.
- [5] Canada Criminal Code. n.d. Section 318-319.
- [6] Rui Cao, Roy Ka-Wei Lee, and Tuan-Anh Hoang. 2020. DeepHate: Hate speech detection via multi-faceted text representations. In *Proceedings of the 12th ACM Conference on Web Science*. 11–20.
- [7] Delan Devakumar, Geordan Shannon, Sunil S. Bhopal, and Ibrahim Abubakar. 2020. Racism and discrimination in COVID-19 responses. *Lancet* 395, 10231 (2020), 1194.
- [8] Facebook. n.d. Hate Speech. Retrieved February 11, 2020 from https://www.facebook.com/communitystandards/hate_speech.
- [9] Jason Faulkner, Mark Schaller, Justin H. Park, and Lesley A. Duncan. 2004. Evolved disease-avoidance mechanisms and contemporary xenophobic attitudes. *Group Processes & Intergroup Relations* 7, 4 (2004), 333–353.
- [10] Paula Fortuna and Sérgio Nunes. 2018. A survey on automatic detection of hate speech in text. *ACM Computing Surveys* 51, 4 (2018), 1–30.
- [11] Björn Gambäck and Utpal Kumar Sikdar. 2017. Using convolutional neural networks to classify hate-speech. In *Proceedings of the 1st Workshop on Abusive Language Online*. 85–90.
- [12] Cherian George. 2015. Hate speech law and policy. In *International Encyclopedia of Digital Communication and Society*. Wiley Blackwell-ICA, 1–10.
- [13] Jeffrey W. Howard. 2019. Free speech and hate speech. *Annual Review of Political Science* 22 (2019), 93–109.
- [14] Ellie Keen, Mara Georgescu, and Rui Gomes. 2016. *Bookmarks: Manual for Combating Hate Speech Online Through Human Rights Education*. COE.
- [15] Xiaofeng Liao and Zhiming Zhao. 2019. Unsupervised approaches for textual semantic annotation, a survey. *ACM Computing Surveys* 52, 4 (2019), 1–45.
- [16] Howard Markel. 1999. When germs travel. *American Scholar* 68, 2 (1999), 61–69.

- [17] Binny Mathew, Ritam Dutt, Pawan Goyal, and Animesh Mukherjee. 2019. Spread of hate speech in online social media. In *Proceedings of the 10th ACM Conference on Web Science*. 173–182.
- [18] Puja Mehta, Daniel F. McAuley, Michael Brown, Emilie Sanchez, Rachel S. Tattersall, and Jessica J. Manson. 2020. COVID-19: Consider cytokine storm syndromes and immunosuppression. *Lancet* 395, 10229 (2020), 1033–1034.
- [19] United Nations. n.d. COVID-19 Stoking Xenophobia, Hate and Exclusion, Minority Rights Expert Warns. Retrieved July 14, 2020 from <https://news.un.org/en/story/2020/03/1060602>.
- [20] United Nations. n.d. COVID-19: UN Counters Pandemic-Related Hate and Xenophobia. Retrieved July 14, 2020 from <https://www.un.org/en/coronavirus/covid-19-un-counters-pandemic-related-hate-and-xenophobia>.
- [21] United Nations. n.d. United Nations Guidance Note on Addressing and Countering COVID-19 Related Hate Speech. Retrieved July 14, 2020 from <https://www.un.org/en/genocideprevention/documents/Guidance%20on%20COVID-19%20related%20Hate%20Speech.pdf>.
- [22] Kunal Relia, Zhengyi Li, Stephanie H. Cook, and Rumi Chunara. 2019. Race, ethnicity and national origin-based discrimination in social media and hate crimes across 100 US cities. In *Proceedings of the International AAAI Conference on Web and Social Media*, Vol. 13. 417–427.
- [23] Leonard Schild, Chen Ling, Jeremy Blackburn, Gianluca Stringhini, Yang Zhang, and Savvas Zannettou. 2020. “Go eat a bat, Chang!”: An early look on the emergence of sinophobic behavior on web communities in the face of COVID-19. arXiv:2004.04046
- [24] Anna Schmidt and Michael Wiegand. 2017. A survey on hate speech detection using natural language processing. In *Proceedings of the 5th International Workshop on Natural Language Processing for Social Media*. 1–10.
- [25] Egyptian Streets. 2020. In the Wake of COVID-19, Egypt’s Asians Fall Victim to Racism. Retrieved July 16, 2020 from <https://egyptianstreets.com/2020/04/15/in-the-wake-of-covid-19-egypts-asians-fall-victim-to-racism/>.
- [26] Times. 2019. Facebook Says It’s Removing More Hate Speech Than Ever Before. But There’s a Catch. Retrieved February 11, 2020 from <https://time.com/5739688/facebook-hate-speech-languages/>.
- [27] Egypt Today. 2020. One of 3 Men Who Bullied Chinese Person as ‘Corona’ Arrested. Retrieved July 16, 2020 from <https://www.egypttoday.com/Article/1/82524/One-of-3-men-who-bullied-Chinese-person-as-corona>.
- [28] Twitter. n.d. Hateful Conduct Policy. Retrieved February 11, 2020 from <https://help.twitter.com/en/rules-and-policies/hateful-conduct-policy>.
- [29] Zeerak Waseem and Dirk Hovy. 2016. Hateful symbols or hateful people? Predictive features for hate speech detection on Twitter. In *Proceedings of the NAACL Student Research Workshop*. 88–93.
- [30] Matthew Williams. 2019. The Connection Between Online Hate Speech and Real-World Hate Crime. Retrieved February 11, 2020 from <https://blog.oup.com/2019/10/connection-between-online-hate-speech-real-world-hate-crime/>.
- [31] Savvas Zannettou, Jason Baumgartner, Joel Finkelstein, Alex Goldenberg, John Farmer, John K. Donohue, and Paul Goldenberg. 2020. *Weaponized Information Outbreak: A Case Study on COVID-19, Bioweapon Myths, and the Asian Conspiracy Meme*. Technical Report. Network Contagion Research Institute.
- [32] Caleb Ziems, Bing He, Sandeep Soni, and Srijan Kumar. 2020. Racism is a virus: Anti-Asian hate and counterhate in social media during the COVID-19 crisis. arXiv:2005.12423

Received July 2020; revised August 2020; accepted October 2020