

# Permanent Twitter ban of extremist influencers can detoxify social media

October 21 2021

---



Credit: CC0 Public Domain

Banning right-wing extremists from social media can reduce the spread of anti-social ideas and conspiracy theories, according to Rutgers-led research.

The study, published in the journal *Proceedings of the ACM on Human-Computer Interaction*, examined what happens after individual

influencers with large followings are banned from social media and no longer have a platform to promote their [extreme views](#).

"Removing someone from a platform is an extreme step that should not be taken lightly," said lead author Shagun Jhaver, an assistant professor in the Department of Library and Information Science at Rutgers-New Brunswick. "However, platforms have rules for appropriate behavior, and when a site member breaks those rules repeatedly, the platform needs to take action. The toxicity created by influencers and their supporters who promote offensive speech can also silence and harm vulnerable user groups, making it crucial for platforms to attend to such influencers' activities."

The study examined three extremist influencers banned on Twitter: Alex Jones, an American radio host and political extremist who gained notoriety for promoting [conspiracy theories](#); Milo Yiannopoulos, a British political commentator who became known for ridiculing Islam, feminism and [social justice](#); and Owen Benjamin, an American "alt-right" actor, comedian and political commentator who promoted anti-Semitic conspiracy theories and anti-LGBT views.

The researchers analyzed more than 49 million tweets referencing the banned influencers, tweets referencing their offensive ideas, and all tweets posted by their supporters six months before and after they were removed from the platform.

Once they were denied social media access, posts referencing each influencer declined by almost 92 percent. The number of existing users and new users specifically tweeting about each influencer also shrank significantly, by about 90 percent.

The bans also significantly reduced the overall posting activity and toxicity levels of supporters. On average, the number of tweets posted by

supporters reduced by 12.59 percent and their toxicity declined by 5.84 percent. This suggests that de-platforming can improve the content quality on the platform.

Researchers say the study indicates that banning those with extremist views who are promoting conspiracy theories minimizes contentious conversations by their supporters. The data from the study will help social media platforms make more informed decisions about whether and when to implement bans, which has been on the rise as a moderation strategy.

"Many people continue to raise concerns about the financial benefits from advertising dollars tied to content that spreads misinformation or conducts harassment," said Jhaver. "This is an opportunity for platforms to clarify their commitment to its users and [de-platform](#) when appropriate. Judiciously using this strategy will allow platforms to address the problem of online radicalization, a worthy goal to pursue even if it leads to short-term loss in advertising dollars."

Future research is needed to examine the interactions between online speech, de-platforming and radicalization and to identify when it would be appropriate to ban users from [social media](#) sites.

**More information:** Shagun Jhaver et al, Evaluating the Effectiveness of Deplatforming as a Moderation Strategy on Twitter, *Proceedings of the ACM on Human-Computer Interaction* (2021). [DOI: 10.1145/3479525](#)

Provided by Rutgers University

Citation: Permanent Twitter ban of extremist influencers can detoxify social media (2021,

October 21) retrieved 7 August 2024 from <https://techxplore.com/news/2021-10-permanent-twitter-extremist-detoxify-social.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.