Twitter was at the forefront of content moderation. What comes next?

April 27 2022, by Suhauna Hussain and Brian Contreras

Credit: Pixabay/CC0 Public Domain

In 2015, a man wearing a skull mask posted a video outlining his plans to murder Brianna Wu. The skull video was only one of many such disturbing and bizarre posts targeting Wu and other women online as part of a harassment campaign dubbed GamerGate.

This was in the early days of content moderation in a social media

ecosystem with far fewer rules, but still, Twitter moved quickly to remove the video, preventing it from going viral, Wu said. Though GamerGate illustrated just how inept social media platforms were at protecting their users, Twitter's prompt action served as an early example of the company's relative willingness to address criticism and work to tamp down on abuse, Wu said.

Although social media platforms have struggled to respond to misinformation, hate speech, election interference and incitement to violence, Twitter has over the years taken a more nuanced and thorough approach, developing, revising and expanding an extensive policy framework.

Twitter, for instance, led efforts to create safety policies and enforce high-profile violations of its rules. It permanently suspended right-wing provocateur Milo Yiannopoulos in July 2016 and conspiracy theorist Alex Jones in September 2018. Facebook did not ban Yiannopoulos and Jones until May 2019.

In summer 2020, Twitter slapped a warning label over then-President Trump's tweet threatening a harsh crackdown on protests in Minneapolis as violating its rules about "glorifying violence," and soon after flagged two more election-related tweets for fact-checking. The move distinguished Twitter from Facebook, whose chief executive, Mark Zuckerberg, indicated it was not inclined to similarly take action, and paved the way for a slew of social media companies that later suspended Trump from their platforms days before the end of his term.

This year, Facebook announced a 24-hour suspension of the account of Rep. Marjorie Taylor Greene, R-Ga., one day after Twitter permanently banned Greene for repeatedly spreading misinformation about COVID-19.

But Elon Musk's successful bid to buy Twitter could change the company's trajectory. Musk, who has said he embraces a philosophy of free speech absolutism, has made it clear he wants a platform with less enforcement, writing in a series of tweets Tuesday that he favors moderation only when legally required.

"I am against censorship that goes far beyond the law. If people want less free speech, they will ask government to pass laws to that effect. Therefore, going beyond the law is contrary to the will of the people," Musk said on Twitter.

"Twitter has historically served as one of the more forward-thinking social media platforms that has always tested new ideas and concepts," said Jennifer Edwards, executive director of the Texas Social Media Research Institute at Tarleton State University.

Musk's purchase of the company and the pivot toward the laissez faire moderation ethos he favors might push other social media platforms to backtrack as well and relax their moderation standards, she said.

The dynamic of Twitter leading the pack could also flow in the opposite direction under Musk, with the new Twitter owner taking cues from his more veteran counterpart over at Facebook. For instance, Musk has already said he wants to begin "authenticating all humans" on Twitter—a move that, though vague, could align his platform more with Zuckerberg's, where users are expected to post under whatever name "they go by in everyday life.

Musk's intimations that he may follow suit have already prompted criticism.

"Any free speech advocate (as Musk appears to view himself) willing to require users to submit ID to access a platform is likely unaware of the

crucial importance of pseudonymity and anonymity," several leaders from the digital rights nonprofit Electronic Frontier Foundation wrote in response to the news of Musk's purchase. Policies that require real names on Facebook have been used to push out precarious communities such as transgender people, drag queens and sex workers, the statement said.

A lot of users understandably distrust social media, don't want to provide identification and drop out, said Sophie Zhang, a former data scientist at Facebook. In South Korea, databases of real name authentication were repeatedly hacked as they formed a treasure trove of personal information, she said.

"Free speech absolutism is a nice idea," but the vast majority of content moderation isn't controversial political discussion Musk posits it to be, and so these values don't necessarily work in practice, Zhang said.

Zhang said it's too early to know how Musk's influence will affect content moderation on the platform. The challenges of the platform may force him and other supporters of free speech absolutism to reckon with the question of why he is unable to let speech flow uninfringed and simultaneously prevent the platform from becoming a morass of crypto spam, pornography and fake adverts.

"The true question to me is how Elon makes those decisions once he is actually in the position of responsibility," she said.

Christopher Bail, a professor at Duke University and director of the campus' Polarization Lab, said the premises of some of Musk's proposals are flawed. Musk is adamant that conservative voices are being minimized, and although it's possible to point to high-profile cases such as the suspension of Trump's account as examples of bias against conservatives, studies show the platform actually tends to promote

conservative perspectives, Bail said.

Musk has said accounts should almost never be banned, but also has promised to crack down on spammers—presumably identifying them by the content of their speech and taking action to ban the accounts.

"I think where the rubber hits the road, it'll be more difficult than he realizes to do what he wants," Bail said.

Researchers and activists worry that Musk's focus on unfettered speech will erode tools Twitter's trust and safety team has built over the years. In giving itself approaches besides account deactivation and post removal, Twitter has made its rules substantially more enforceable, experts said. The company has exercised greater transparency than its peers, maintaining open lines of communication with researchers and made swaths of data around spam and misinformation on the platform public and available for analysis by academics and others.

Twitter maintains an archive of posts it has removed from the platform, allowing researchers to examine the reach and influence of viral misinformation. Twitter's Birdwatch initiative aims to create a crowdsourced approach to flagging of misinformation.

Wu said that during the peak of harassment she faced during GamerGate, Twitter's then vice president of trust and safety reached out to hear concerns and offer support. In the years following, Twitter separated itself from the pack by making real efforts to engage with critics like herself.

"They did more than Facebook did, more than Reddit, more than Google," said Wu, who says she informally advised the company's trust and safety team in an unpaid capacity for about five years. "Twitter has never gotten the credit it deserves for addressing harassment

aggressively."

©2022 Los Angeles Times.
Distributed by Tribune Content Agency, LLC.

Citation: Twitter was at the forefront of content moderation. What comes next? (2022, April 27) retrieved 27 April 2024 from
https://techxplore.com/news/2022-04-twitter-forefront-content-moderation.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.