

Moderating online content increases accountability, but can harm some platform users

July 27 2022



Credit: Pixabay/CC0 Public Domain

Marginalized social media users face disproportionate content removal from platforms, but the visibility of this online moderation is a double-

edged sword.

A new University of Michigan study about two [online platforms](#)—Reddit and Twitch—suggests that greater visibility means increased accountability, but the actions can also bring more attention to the offensive content, further harming marginalized people.

Content moderation protects [platform](#) users from each other by removing offensive or illegal messages. Despite adhering to site policies and community norms, often times the moderated content targets marginalized groups, such as racial and gender minorities and queer people, said Oliver Haimson, assistant professor at the U-M School of Information and Digital Studies Institute.

"Content moderation is often difficult to study, as its visibility depends on a platform's tools and moderators' decision on whether to make their processes visible or not," said study lead author Hibby Thach, a master's degree student at the University of Illinois Chicago and incoming doctoral student at U-M.

In the study published in *New Media & Society*, the researchers analyzed live chat and streaming in a Twitch category involving women, people of color and LGBTQ+ people, and the text-based Reddit communities of transgender users. The visibility of content moderation processes differed between the two platforms.

On Reddit, content moderation becomes most visible during direct, text-based comment interactions between users and moderators, which generally occur after a moderator removes a user's post or comment. Reddit users challenge the actions of subreddit moderators, through appeals or requests for explanation, when content is removed.

On Twitch, streamers and viewers actively participate in the content

moderation process or discuss content moderation decisions during livestreams, revealing their interventions with the phrase "message deleted by a moderator" and through "unban appeals."

Visibility of moderation for both platforms is needed to address accountability and concerns around unfair content removal or account suspension, the researchers said. However, also in both cases, visibility is less critical in instances where content obviously violates community guidelines, particularly abusive and harmful content, they said.

Media attention focused on Reddit and Twitch has contributed to content moderation visibility. During the study, researchers saw greater transphobic activity on Reddit, as well as misogynistic and racist activity on Twitch. But marginalized users have also received [social support](#). While community moderators have attempted to protect users, the platforms must continue using tools and detection devices to lessen the negative content, Thach said.

The study's other authors are U-M research assistant Samuel Mayworm and U-M doctoral student Daniel Delmonaco.

More information: Hibby Thach et al, (In)visible moderation: A digital ethnography of marginalized users and content moderation on Twitch and Reddit, *New Media & Society* (2022). [DOI: 10.1177/14614448221109804](#)

Provided by University of Michigan

Citation: Moderating online content increases accountability, but can harm some platform users (2022, July 27) retrieved 18 April 2024 from <https://techxplore.com/news/2022-07-moderating-online-content-accountability-platform.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.