

A researcher asked COVID anti-vaxxers how they avoid Facebook moderation. Here's what they found

August 17 2022, by Damilola Ayeni, Belinda Barnet and Diana Bossio



Credit: AI-generated image ([disclaimer](#))

How are social media platforms managing vaccine misinformation at this stage in the pandemic?

Anti-vaccine sentiment has been building since 2020, and hasn't gone

anywhere. In fact, it will have intensified following the recent approval of COVID-19 vaccinations for some [babies and children under five](#), and the recommendation for a fourth booster shot for people [over 30](#).

And although anti-vaxxers can be found in most online spaces, Facebook has historically been one of [their platforms](#) of choice.

Swinburne Ph.D. student Damilola Ayeni has been interviewing anti-vaccine activists since 2019, to learn about how they grow their audience on Facebook and how they evade moderation.

Her findings help shed light on the tug-of-war between [Facebook's](#) content moderation efforts and an unrelenting slew of vaccine misinformation.

What's been happening?

Facebook has been moderating content under the [COVID-19 and vaccine policy](#). It does this by warning group admins and moderators, deleting offending accounts or groups, and flagging posts containing misinformation.

In its first response to Australia's DIGI Misinformation and Disinformation Code, Facebook said it had "removed over 14 million pieces of content that constituted misinformation related [to COVID-19](#)"—of which 110,000 were from Australian pages.

Despite this, Facebook's moderation approach has loopholes that anti-vaxxers continue to exploit. For instance, the ABC recently [fact-checked](#) anti-vaxxers who were spreading misinformation on Facebook by claiming COVID-19 vaccines were responsible for the sudden death of a Queensland toddler.

Ayeni's research found anti-vax Facebook groups are now "self-moderating." This means they predict what Facebook's automated moderation tools and independent fact-checkers will be looking for, and change their posting techniques accordingly.

Group members share in-house "rules" to help guide content strategies. In some cases, group administrators will allow content to stay up for a short time, so there's opportunity to see it before it's flagged by Facebook.

One anti-vaxxer told Ayeni they now conduct more research on other members' posts; if the content is obviously untrue or controversial, they delete the post themselves.

Ayeni also found content that's likely to be targeted by fact-checkers or automated moderation is creatively manipulated. For instance, users may use screenshots or images to avoid text-based moderation. Or they may intentionally misspell key words such as "anti-vaccine," or leave them out altogether.

Satire and sarcasm are also used in an effort to misdirect Facebook's fact-checkers, while "signaling" the poster's vaccine beliefs to like-minded users. One post seen by Ayeni sarcastically challenged the government to get a "real" COVID-19 vaccine before administering it to the public.

Are anti-vaxxers moving away from Facebook?

Interviewees said they initially gravitated towards Facebook because it met some of their privacy needs, including the ability to create private and secret groups.

In 2019, Facebook began a platform redesign focused on improving users' privacy. Its [goal was](#) to encourage more encrypted and intimate

forms of communication through Messenger and in Facebook groups. And this brought along features that attracted anti-vaxxers to the platform early in the pandemic.

Moderators told Ayeni Facebook groups provided an environment where they could safely offer support to other members and build communities for "like-minded" individuals.

However, Facebook's increased moderation has undoubtedly made it less attractive. Some users said they want to leave altogether due to the consistent [reporting of their accounts](#) and difficulty fighting [platform](#) decisions.

Many were looking to migrate to less moderated platforms such as Telegram, Parler, MeWe, Mighty Networks and Wimkim. All of these are uncensored, unmoderated and all too easy to access.

Telegram in particular is now favored by [far-right and conspiracy groups](#). It has also attracted high-profile anti-vaxxers including former TV presenter Pete Evans and former Liberal MPs George Christensen and Craig Kelly—individuals who were repeatedly moderated and eventually de-platformed from [Facebook's products](#).

In April 2021, Facebook banned Kelly for breaching its misinformation policies in relation to COVID-19 and vaccinations. At the time he claimed Facebook "[burnt and torched and incinerated](#)" his voice, but his following on Telegram has swelled from 10,000 back then to about 74,000 now.

What can be done?

Facebook has become increasingly reliant on automated moderation during the pandemic. This experiment has [not gone well](#) for it. Machine-

learning algorithms still can't detect wordplay, sarcasm, and embedded messaging in images as well as human moderators can.

We believe platforms need to recognize anti-vaxxers' tactics are evolving to keep pace with moderation tools. And meaningful push-back will require more investment in human moderators, not just AI.

At the same time, it would make sense to ensure other platforms operating in Australia—such as Telegram, for instance—are subject to the same regulatory scrutiny as Facebook is. Until these smaller platforms also take responsibility for vaccine misinformation, they will continue to be a magnet for it.

This article is republished from [The Conversation](#) under a Creative Commons license. Read the [original article](#).

Provided by The Conversation

Citation: A researcher asked COVID anti-vaxxers how they avoid Facebook moderation. Here's what they found (2022, August 17) retrieved 24 June 2024 from <https://medicalxpress.com/news/2022-08-covid-anti-vaxxers-facebook-moderation.html>

| |
|--|
| <p>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.</p> |
|--|