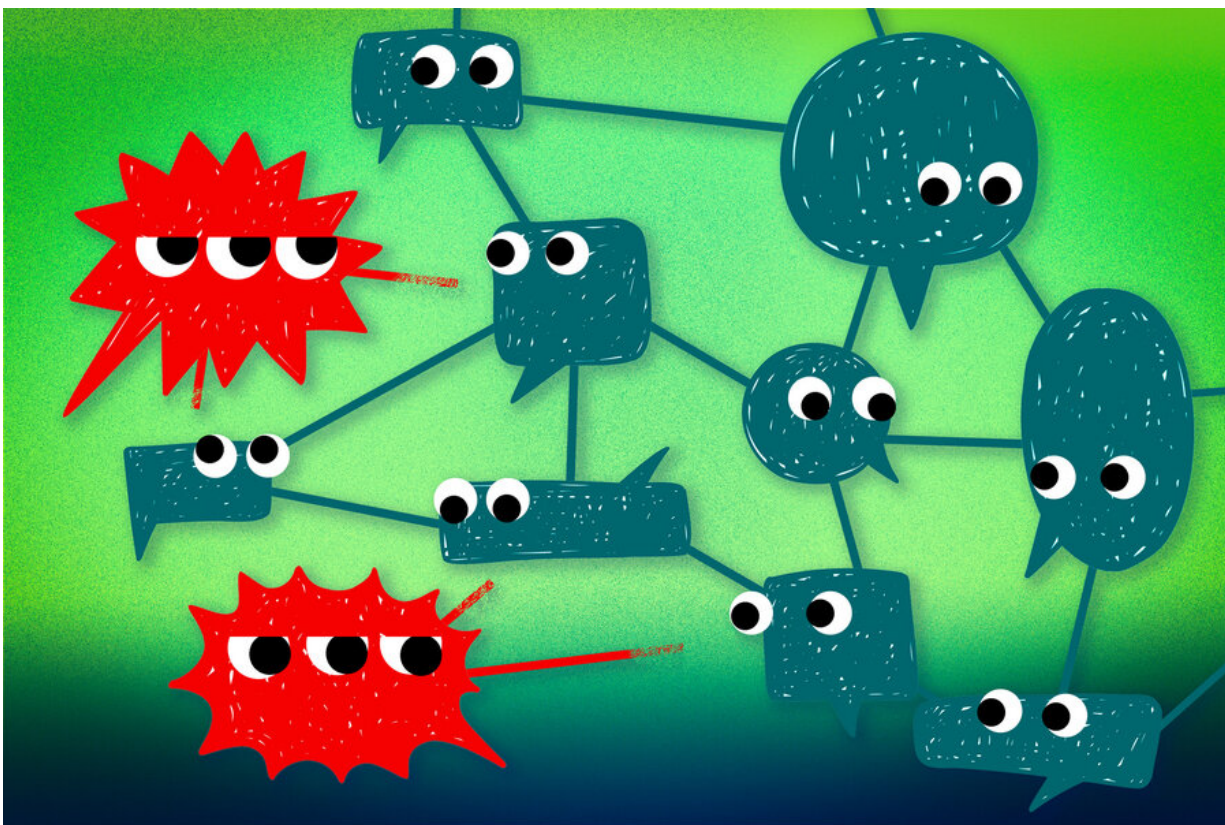


# Empowering social media users to assess content helps fight misinformation

November 16 2022, by Adam Zewe

---



MIT researchers built a prototype social media platform to study the effects of giving users more agency to assess content for accuracy and control the posts they see based on accuracy assessments from others. Credit: Image: Jose-Luis Olivares, MIT

When fighting the spread of misinformation, social media platforms

typically place most users in the passenger seat. Platforms often use machine-learning algorithms or human fact-checkers to flag false or misinforming content for users.

"Just because this is the status quo doesn't mean it is the correct way or the only way to do it," says Farnaz Jahanbakhsh, a graduate student in MIT's Computer Science and Artificial Intelligence Laboratory (CSAIL).

She and her collaborators conducted a study in which they put that power into the hands of social media users instead.

They first surveyed people to learn how they avoid or filter misinformation on social media. Using their findings, the researchers developed a prototype [platform](#) that enables users to assess the accuracy of content, indicate which users they trust to assess accuracy, and filter posts that appear in their feed based on those assessments.

Through a [field study](#), they found that users were able to effectively assess misinforming posts without receiving any prior training. Moreover, users valued the ability to assess posts and view assessments in a structured way. The researchers also saw that participants used content filters differently—for instance, some blocked all misinforming content while others used filters to seek out such articles.

This work shows that a decentralized approach to moderation can lead to higher content reliability on social media, says Jahanbakhsh. This approach is also more efficient and scalable than centralized moderation schemes, and may appeal to users who mistrust platforms, she adds.

"A lot of research into misinformation assumes that users can't decide what is true and what is not, and so we have to help them. We didn't see that at all. We saw that people actually do treat content with scrutiny and

they also try to help each other. But these efforts are not currently supported by the platforms," she says.

Jahanbakhsh wrote the paper with Amy Zhang, assistant professor at the University of Washington Allen School of Computer Science and Engineering; and senior author David Karger, professor of computer science in CSAIL. The research will be presented at the ACM Conference on Computer-Supported Cooperative Work and Social Computing, and is published as part of the *Proceedings of the ACM on Human-Computer Interaction*.

## **Fighting misinformation**

The spread of online misinformation is a widespread problem. However, current methods social media platforms use to mark or remove misinforming content have downsides. For instance, when platforms use algorithms or fact-checkers to assess posts, that can create tension among users who interpret those efforts as infringing on freedom of speech, among other issues.

"Sometimes users want misinformation to appear in their feed because they want to know what their friends or family are exposed to, so they know when and how to talk to them about it," Jahanbakhsh adds.

Users often try to assess and flag misinformation on their own, and they attempt to assist each other by asking friends and experts to help them make sense of what they are reading. But these efforts can backfire because they aren't supported by platforms. A user can leave a comment on a misleading post or react with an angry emoji, but most platforms consider those actions signs of engagement. On Facebook, for instance, that might mean the misinforming content would be shown to more people, including the user's friends and followers—the exact opposite of what this user wanted.

To overcome these problems and pitfalls, the researchers sought to create a platform that gives users the ability to provide and view structured accuracy assessments on posts, indicate others they trust to assess posts, and use filters to control the content displayed in their feed. Ultimately, the researchers' goal is to make it easier for users to help each other assess misinformation on social media, which reduces the workload for everyone.

The researchers began by surveying 192 people, recruited using Facebook and a mailing list, to see whether users would value these features. The survey revealed that users are hyper-aware of misinformation and try to track and report it, but fear their assessments could be misinterpreted. They are skeptical of platforms' efforts to assess content for them. And, while they would like filters that block unreliable content, they would not trust filters operated by a platform.

Using these insights, the researchers built a Facebook-like prototype platform, called Trustnet. In Trustnet, users post and share actual, full news articles and can follow one another to see content others post. But before a user can post any content in Trustnet, they must rate that content as accurate or inaccurate, or inquire about its veracity, which will be visible to others.

"The reason people share misinformation is usually not because they don't know what is true and what is false. Rather, at the time of sharing, their attention is misdirected to other things. If you ask them to assess the content before sharing it, it helps them to be more discerning," she says.

Users can also select trusted individuals whose content assessments they will see. They do this in a private way, in case they follow someone they are connected to socially (perhaps a friend or family member) but whom they would not trust to assess content. The platform also offers filters

that let users configure their feed based on how posts have been assessed and by whom.

## Testing Trustnet

Once the prototype was complete, they conducted a study in which 14 individuals used the platform for one week. The researchers found that users could effectively assess content, often based on expertise, the content's source, or by evaluating the logic of an article, despite receiving no training. They were also able to use filters to manage their feeds, though they utilized the filters differently.

"Even in such a small sample, it was interesting to see that not everybody wanted to read their news the same way. Sometimes people wanted to have misinforming posts in their feeds because they saw benefits to it. This points to the fact that this agency is now missing from [social media](#) platforms, and it should be given back to users," she says.

Users did sometimes struggle to assess content when it contained multiple claims, some true and some false, or if a headline and article were disjointed. This shows the need to give users more assessment options—perhaps by stating that an article is true-but-misleading or that it contains a political slant, she says.

Since Trustnet users sometimes struggled to assess articles in which the content did not match the headline, Jahanbakhsh launched [another research project to create a browser extension](#) that lets users modify news headlines to be more aligned with the article's content.

While these results show that users can play a more active role in the fight against misinformation, Jahanbakhsh warns that giving users this power is not a panacea. For one, this approach could create situations where users only see information from like-minded sources. However,

filters and structured assessments could be reconfigured to help mitigate that issue, she says.

In addition to exploring Trustnet enhancements, Jahanbakhsh wants to study methods that could encourage people to read content assessments from those with differing viewpoints, perhaps through gamification. And because [social media platforms](#) may be reluctant to make changes, she is also developing techniques that enable [users](#) to post and view content assessments through normal web browsing, instead of on a platform.

**More information:** Farnaz Jahanbakhsh et al, Leveraging Structured Trusted-Peer Assessments to Combat Misinformation, *Proceedings of the ACM on Human-Computer Interaction* (2022). [DOI: 10.1145/3555637](#)

Farnaz Jahanbakhsh et al, Our Browser Extension Lets Readers Change the Headlines on News Articles, and You Won't Believe What They Did!, *Proceedings of the ACM on Human-Computer Interaction* (2022). DOI: 10.1145/3555643 , [dl.acm.org/doi/10.1145/3555643](https://dl.acm.org/doi/10.1145/3555643)

Provided by Massachusetts Institute of Technology

Citation: Empowering social media users to assess content helps fight misinformation (2022, November 16) retrieved 17 April 2024 from <https://techxplore.com/news/2022-11-empowering-social-media-users-content.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.