New browser extension empowers users to fight online misinformation

May 16 2024, by Adam Zewe

Outgoing links to a Youtube video on the page are displayed with a checkmark icon (marked with blue arrows), indicating that the video has been assessed as accurate by the user's assessors. Credit: *arXiv* (2024). DOI: 10.48550/arxiv.2403.11485

Most people agree that the spread of online misinformation is a serious problem. But there is much less consensus on what to do about it.

Many proposed solutions focus on how social media platforms can or should moderate content their users post, to prevent misinformation from spreading.

"But this approach puts a critical social decision in the hands of for-profit companies. It limits the ability of users to decide who they trust. And having platforms in charge does nothing to combat misinformation users come across from other online sources," says Farnaz Jahanbakhsh SM '21, Ph.D. '23, who is currently a postdoc at Stanford University.

She and MIT Professor David Karger have proposed an alternate strategy. They built a web browser extension that empowers individuals to flag misinformation and identify others they trust to assess online content.

Their decentralized approach, called the Trustnet browser extension, puts the power to decide what constitutes misinformation into the hands of individual users rather than a central authority. Importantly, the universal browser extension works for any content on any website, including posts on social media sites, articles on news aggregators, and videos on streaming platforms.

Through a two-week study, the researchers found that untrained individuals could use the tool to effectively assess misinformation. Participants said having the ability to assess content, and see assessments from others they trust, helped them think critically about it.

"In today's world, it's trivial for bad actors to create unlimited amounts of misinformation that looks accurate, well-sourced, and carefully argued. The only way to protect ourselves from this flood will be to rely on information that has been verified by trustworthy sources. Trustnet presents a vision of how that future could look," says Karger.

Jahanbakhsh, who conducted this research while she was an [electrical engineering](link) and computer science (EECS) graduate student at MIT, and Karger, a professor of EECS and a member of the Computer Science and Artificial Intelligence Laboratory (CSAIL), detail their findings in a

paper presented this week at the ACM Conference on Human Factors in Computing Systems ([CHI 2024](link)). The [research](link) is also available on the *arXiv* preprint server.

Fighting misinformation

This new paper builds off their [prior work](link) about fighting online misinformation. The researchers built a social media platform called Trustnet, which enabled users to assess content accuracy and specify trusted users whose assessments they want to see.

But in the real world, few people would likely migrate to a new social media platform, especially when they already have friends and followers on other platforms. On the other hand, calling on social media companies to give users content-assessment abilities would be an uphill battle that may require legislation. Even if regulations existed, they would do little to stop misinformation elsewhere on the web.

Instead, the researchers sought a platform-agnostic solution, which led them to build the Trustnet browser extension.

Extension users click a button to assess content, which opens a side panel where they label it as accurate, inaccurate, or question its accuracy. They can provide details or explain their rationale in an accompanying text box.

Users can also identify others they trust to provide assessments. Then, when the user visits a website that contains assessments from these trusted sources, the side panel automatically pops up to show them.

In addition, users can choose to follow others beyond their trusted assessors. They can opt to see content assessments from those they follow on a case-by-case basis. They can also use the side panel to

respond to questions about content accuracy.

"But most content we come across on the web is embedded in a social media feed or shown as a link on an aggregator page, like the front page of a news website. Plus, something we know from prior work is that users typically don't even click on links when they share them," Jahanbakhsh says.

To get around those issues, the researchers designed the Trustnet Extension to check all links on the page a user is reading. If trusted sources have assessed content on any linked pages, the extension places indictors next to those links and will fade the text of links to content deemed inaccurate.

One of the biggest technical challenges the researchers faced was enabling the link-checking functionality since links typically go through multiple redirections. They were also challenged to make design decisions that would suit a variety of users.

Differing assessments

To see how individuals would utilize the Trustnet Extension, they conducted a two-week study where 32 individuals were tasked with assessing two pieces of content per day.

The researchers were surprised to see that the content these untrained users chose to assess, such as home improvement tips or celebrity gossip, was often different from content assessed by professionals, like news articles. Users also said they would value assessments from people who were not professional fact-checkers, such as having doctors assess medical content or immigrants assess content related to foreign affairs.

"I think this shows that what users need and the kinds of content they

consider important to assess doesn't exactly align with what is being delivered to them. A decentralized approach is more scalable, so more content could be assessed," Jahanbakhsh says.

However, the researchers caution that letting users choose whom to trust could cause them to become trapped in their own bubble and only see content that agrees with their views.

This issue could be mitigated by identifying trust relationships in a more structured way, perhaps by suggesting a user follow certain trusted assessors, like the FDA.

In the future, Jahanbakhsh wants to further study structured trust relationships and the broader implications of decentralizing the fight against misinformation. She also wants to extend this framework beyond misinformation. For instance, one could use the tool to filter out content that is not sympathetic to a certain protected group.

"Less attention has been paid to decentralized approaches because some people think individuals can't assess content," she says. "Our studies have shown that is not true. But users shouldn't just be left helpless to figure things out on their own. We can make fact-checking available to them, but in a way that lets them choose the content they want to see."