# *Putting Users in Control of Social Platforms for Better Content Credibility*

# Putting Users in Control of Social Platforms for Better Content Credibility

Farnaz Jahanbakhsh
Computer Science and Artificial Intelligence Laboratory, Massachusetts Institute of Technology
Cambridge, USA

**ACM Reference Format:**
Farnaz Jahanbakhsh. 2022. Putting Users in Control of Social Platforms for Better Content Credibility. In *Computer Supported Cooperative Work and Social Computing (CSCW'22 Companion), November 8–22, 2022, Virtual Event, Taiwan.* ACM, New York, NY, USA, 4 pages. https://doi.org/10.1145/3500868.3561399

My research is on re-imagining the design of social media and more broadly the web to achieve greater end user empowerment. I believe one of the implications of this greater user empowerment can be higher credibility of content. In this pursuit, I investigate alternative designs to the status quo of the existing social media platforms which prioritize engagement over accuracy.

It is known that misinformation—especially that which promotes outrage—can generate substantial engagement [3]. Therefore, platforms, despite their visible efforts to combat misinformation, may in fact be reluctant to succeed in increasing accuracy. In fact, their prioritization of engagement over accuracy can at times undermine users' own efforts to combat misinformation on the platforms. For instance, a user who encounters an incorrect post and then leaves a comment refuting it may in fact be inadvertently disseminating it farther because the system considers the comment as engagement. Another related failure in design is the lack of clear signals for capturing and displaying content credibility, leading engagement metrics such as the number of likes and shares to affect users' perception of content credibility [2]. Yet another suboptimal design decision is an emphasis on low barriers to sharing that allows users to share content without much attention to its accuracy or potential negative consequences, simply to receive social feedback or elevated engagement [4, 7].

## EMPOWERING PEOPLE TO NAVIGATE THEIR SOCIAL INFORMATION SPACE

In re-imagining the design of social content sharing platforms, my goal is not to impose on users what they should or should not see, but rather give them the right tools that they can use to determine the credibility of content—not only for themselves, but also for the benefit of their social circle. Some of my research tools additionally give people the ability to modify content on the web for the better and be authors of truth in their own right.

In my work, I have explored how the share procedure on social media could be altered to nudge users to have accuracy on top of their mind. Through a collaboration with researchers in Political Science, Psychology, Management Science, and Computer Science, I designed interventions that could be deployed on platforms at scale and which proved to be effective at reducing the sharing of misinformation. These interventions include (i) requiring sharers to click a button to indicate whether they think a post is accurate or not when sharing it, (ii) requiring sharers to choose at least one tag from a small checklist indicating why they consider it accurate, and (iii) writing a short comment explaining their accuracy assessment. As part of designing these interventions, I developed a taxonomy of reasons why people believe or disbelieve news claims [8]. The purpose of the taxonomy was for its categories to be presented in a checklist form so that users could select their reasoning from the set. The attempt to direct users' attention to accuracy is tied to prior work which argues that although people are generally good at discerning accuracy, when deciding whether to share content, they are less concerned with accuracy than other aspects of sharing such as the amount of social feedback they receive [12, 13]. The assessments provided by sharers could also provide valuable information to other users seeking to form their own judgment of a post's accuracy.

I have also explored user affordances which, if deployed on social media, can give social media users greater power in their fight against misinformation. These affordances include (i) the ability to provide structured accuracy assessments of posts and for the assessments to be captured as part of the data-model, (ii) user-specified indication of trust in other users, and (iii) providing filters that users can configure to block posts from their feed based on the accuracy status of posts assessed by the users' trusted sources.

To determine whether these ideas might address users' current needs, I surveyed a diverse group of people about their practices reading and sharing online content and what accuracy-oriented features they would like to have in a news reading and sharing platform.

I found that people already produce and consume accuracy assessments, and ask their social circle and are asked about the accuracy of news posts. To do so, they repurpose the features that are offered by platforms, such as comments and likes that can unfortunately be picked up by platforms as signs of engagement. The survey informed us that users have trusted sources and are trusted by others to relay fact-checking information; and there is enthusiasm for using such assessments in filtering, as long as the filtering remains under user control. To test how users might actually use these new affordances if given the opportunity, I built and conducted a user study on a social content sharing sharing


*CSCW'22 Companion, November 8–22, 2022, Virtual Event, Taiwan*

ACM ISBN 978-1-4503-9190-0/22/11.
https://doi.org/10.1145/3500868.3561399

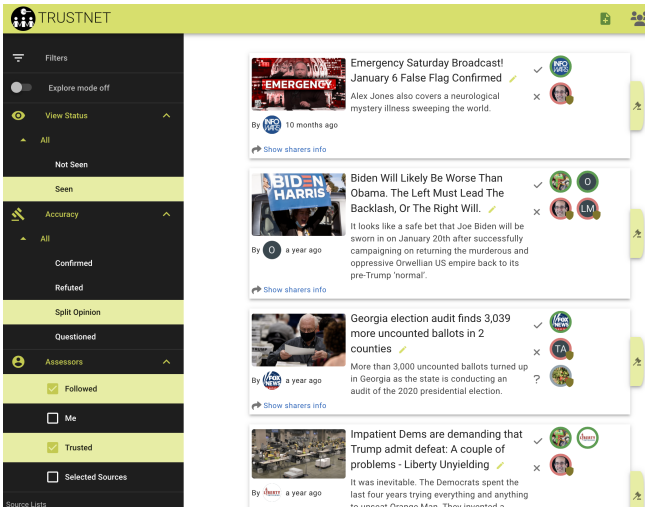


**Figure 1: Users can filter their feed by assessments of their trusted sources**

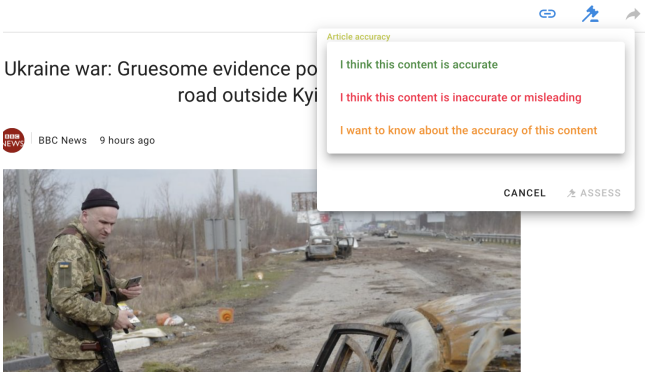


**Figure 2: Users can assess content as (in)accurate and provide their reasoning or ask about its accuracy**

platform prototype[1] that embraces the paradigm of user empowerment described above, where users can assess content, specify the sources they trust, and filter their news feed based on accuracy assessments provided by their trusted sources. The platform displays structured accuracy assessments next to posts (Figures 1 & 2). Study participants used the system as intended and saw value in vetting posts, facilitating inquiries about the accuracy of posts, and seeing assessments from others. The user study also revealed the challenges of the affordances in their current form that need to be addressed going forward. These include the need for separation of various dimensions of credibility when assessing a post, including accuracy, true-but-misleading statements, slant, or a mismatch between content and headline [10]. The next section will contain plans for further exploring the space and implications of these design affordances.

The adoption of these proposed affordances requires persuading platforms, perhaps through activism or legislation, to adopt a

[1]https://trustnet.csail.mit.edu

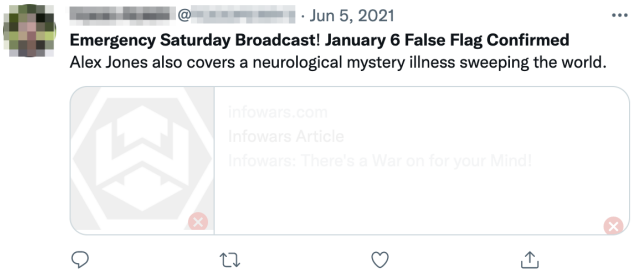


**Figure 3: The Trustnet extension. A link assessed as inaccurate has an icon next to it and is faded.**

content curation model that is not as focused on increasing user engagement as the present. Until that goal is realized, we need to explore ways to leverage the affordances without compliance from the platforms. In pursuing this goal, I have built a browser extension that allows users to assess any page on the web as accurate or inaccurate or they can use it to inquire about content accuracy. The extension in essence provides an overlay on the web and can be used to assess any content with a unique URL including new articles, tweets, Facebook memes, Youtube videos, etc. without involvement from the underlying platforms. When the followers of a user who has assessed a page open the page, they will see the user's assessments on the page (Figure 4). However, because links on social media are often not clicked on, even when they are shared [6], showing the assessments of a page when a user lands on the page will not benefit the multitude of those users who do not navigate to the page. To address this issue, the extensions looks for all the links on the page that the user is viewing, and if there are assessments for any of the links, it will indicate the assessed accuracy of those links by placing indicator icons next to them (Figure 3).

The content that has the potential to deceive or mislead users on social platforms comes in many forms. One complex case is that of news headlines. Headlines play a critical role in steering consumers to news. But the interests of headline publishers are not always aligned with those of consumers. While consumers may want to be informed or entertained, publishers may wish to attract clicks to stories to earn ad revenue [11, 14] and malicious actors may wish to disinform users or manipulate their opinions. The manipulation techniques, employed by disinformation news websites and also to some extent by legitimate sources, include using language that is exaggerated, sensationalized, teasing, misleading, or even inaccurate [5]. With a large portion of news consumption these days happening on social media streams where headlines are detached from an article's content and context [1, 5], misinforming headlines have the potential to be believed and further spread [5].

To address this issue, I explored a new approach that empowers any user to contribute alternative headlines for news articles. I created a browser extension that identifies an article's headline and allows the user to submit an alternative headline in its place. Other consumers who follow that user will see those alternate headlines *anywhere on the web*—including the homepage of the news website,

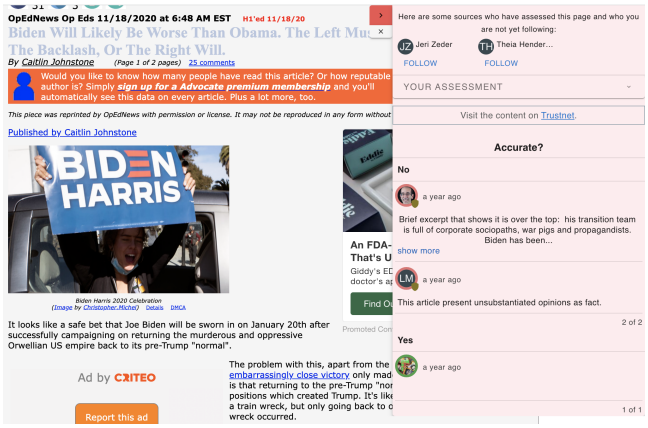


**Figure 4: On any page on the web, the Trustnet extension shows assessments of the users' trusted sources.**


**Figure 5: On any page, the Reheadline extension looks for all headlines for which alternative headlines have been suggested and places the suggested headline next to the original**

social media feeds, or news aggregators—where the original headline appears (Figure 5). Through a user study, we demonstrated that users are both capable of and interested in improving the headlines that they encounter. Our findings suggested that there is potential for including users in the content curation process and not treating them as passive consumers [9].

## RESEARCH AGENDA

The design space for the affordances that can give users more power and control over their social information space is vast; and there are many design decisions left unexplored that I consider exciting future directions for my research.

### Leveraging crowdsourced assessments

The benefit of the design affordance of user-specified trust in other sources to assess content can be constrained under certain circumstances. For instance, it is conceivable that not all users immediately know trusted sources with specialized knowledge and credentials who can assess content on a variety of topics such as climate science or medicine. To address this gap, I aim to explore how trust relationships can be extended to build a trust network for each user who through others may be connected to sources they do not know. In this scenario, when a user leaves a credibility assessment on a post, their assessment can be propagated to all the users that either immediately trust or have an indirect trust path to the assessor, while maintaining the assessor's anonymity (because we have designed a user's trust relationships to be private to the user). This chain of trust could help users benefit from more extensive assessments even though they may not immediately know the benefactors. Transitive trust relationships will allow us to investigate several interesting research questions. One is how fast trust decays as the distance of two sources in the network who are connected by an implicitly inferred trust relationship increases. Another is how assessments from different sources, some not immediately connected to the user, should be weighted, aggregated, and presented to the user in an interprettable manner. One scenario could be that each trusted source is given an equal weight in deciding the accuracy of an article. Conversely, the user could decide that a particular source or rationale be given priority over others. These questions can benefit from both qualitative inquiries as well as field deployments of prototype tools where we can measure aspects of user behavior and the impact of our design decisions quantitatively.

### Towards credibility tools that do not rely on underlying platforms

Some of the tools that I have built in my research (the Trustnet and Reheadline extensions) act as overlays on the entire web so that they can be versatile and platform agnostic. I aim to expand this class of tools that give users more power to help themselves and their social circle determine the credibility of content. For instance, as all media and sources have their biases, any one news story by a source can be slanted or perhaps may not cover the whole story. By providing an overlay tool where anyone can suggest alternative stories or sources on (i.e., in lieu of) any news story on the web, and any user's free choice of whose suggestions they want to heed, users can be better equipped to obtain a more comprehensive and accurate picture of a story.

### Studying and building tools to tackle other forms of misinformation

The affordances I have given users for assessing accuracy of content were mostly developed and tested on factual content. I plan to explore what labels of accuracy may be appropriate for other types of content, such as opinion pieces and satire. This is an important problem because these type of content are often shared on social media where their non-factuality status is obfuscated or ignored by or unknown to users. Additionally, I aim to investigate to what extent sharing of misinformation through ephemeral posting, i.e., stories that disappear after a limited time, is rampant. If ephemeral

misinformation turns out to be a problem, identifying and fact-checking the content as well as notifying users of its accuracy will be so time-sensitive that will require the pipeline to be reconsidered and specifically designed for this domain.

## CONCLUSION

As misinformation continues to impact people's lives and livelihood, it has become increasingly important to understand and help the online information consumption of users. My research involves re-imagining the design of platforms for sharing content and more broadly the web. In doing so, I provide users with tools that they can use to filter out credible from inaccurate content as well as empower them to not be mere passive consumers, and instead do something about the inaccurate or misleading content that they encounter on the web.

## ACKNOWLEDGMENTS

I am grateful to my advisor Professor David Karger for his continuous guidance, support, and feedback on my work. This work was partially supported by NSF award 1915724.

## REFERENCES

[1] Ben Adler. 2014. Tabloids in the age of social media. *Columbia Journalism Review* (2014).

[2] Mihai Avram, Nicholas Micallef, Sameer Patil, and Filippo Menczer. 2020. Exposure to social engagement metrics increases vulnerability to misinformation. *Harvard Kennedy School Misinformation Review* 1 (07 2020). https://doi.org/10.37016/mr-2020-033

[3] Vian Bakir and Andrew McStay. 2018. Fake news and the economy of emotions: Problems, causes, solutions. *Digital journalism* 6, 2 (2018), 154–175.

[4] Natalya N Bazarova, Yoon Hyung Choi, Victoria Schwanda Sosik, Dan Cosley, and Janis Whitlock. 2015. Social sharing of emotions on Facebook: Channel differences, satisfaction, and replies. In *Proceedings of the 18th ACM conference on computer supported cooperative work & social computing*. 154–164.

[5] Yimin Chen, Niall J Conroy, and Victoria L Rubin. 2015. Misleading online content: recognizing clickbait as" false news". In *Proceedings of the 2015 ACM on workshop on multimodal deception detection*. 15–19.

[6] Maksym Gabielkov, Arthi Ramachandran, Augustin Chaintreau, and Arnaud Legout. 2016. Social clicks: What and who gets read on Twitter?. In *Proceedings of the 2016 ACM SIGMETRICS international conference on measurement and modeling of computer science*. 179–192.

[7] Nir Grinberg, Shankar Kalyanaraman, Lada A Adamic, and Mor Naaman. 2017. Understanding feedback expectations on Facebook. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*. 726–739.

[8] Farnaz Jahanbakhsh, Amy X Zhang, Adam J Berinsky, Gordon Pennycook, David G Rand, and David R Karger. 2021. Exploring lightweight interventions at posting time to reduce the sharing of misinformation on social media. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW1 (2021), 1–42.

[9] Farnaz Jahanbakhsh, Amy X Zhang, Karrie Karahalios, and David R Karger. 2022. 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* 6, CSCW2 (2022), 1–34.

[10] Farnaz Jahanbakhsh, Amy X Zhang, and David R Karger. 2022. Leveraging Structured Trusted-Peer Assessments to Combat Misinformation. *Proceedings of the ACM on Human-Computer Interaction* 6, CSCW2 (2022), 1–40.

[11] Terrance McCoy. 2016. Inside a Long Beach Web operation that makes up stories about Trump and Clinton: What they do for clicks and cash. *Los Angeles Times* (2016).

[12] Gordon Pennycook, Ziv Epstein, Mohsen Mosleh, Antonio A Arechar, Dean Eckles, and David G Rand. 2021. Shifting attention to accuracy can reduce misinformation online. *Nature* 592, 7855 (2021), 590–595.

[13] Gordon Pennycook, Jonathon McPhetres, Yunhao Zhang, Jackson G Lu, and David G Rand. 2020. Fighting COVID-19 misinformation on social media: Experimental evidence for a scalable accuracy-nudge intervention. *Psychological science* 31, 7 (2020), 770–780.

[14] Alexander Smith. 2016. Fake News: How a Partying Macedonian Teen Earns Thousands Publishing Lies. *NBC News* (2016).