

News and Perspectives

How to Use—And How Not to Use—Public Forums in Research

Simon Spichak, JMIR Correspondent

Abstract

Public forums can be a boon for researchers—but also present thorny ethical issues. In this *News and Perspectives* article, JMIR Correspondent Simon Spichak reports on some of the challenges, missteps, and necessary corrections for using public forums in research.

Key Takeaways:

- Publicly available comments and data on social media are fair game for researchers to study, even if most posters are unaware.
- Unethical experiments can erode trust in online communities.
- It is possible to run a social media study that benefits users.

On March 17, 2025, the moderators of the /r/ChangeMy-View (CMV) subreddit received an alarming message from University of Zurich researchers. “Over the past few months, we used multiple accounts to post AI-written comments in posts published on CMV,” the message said.

Researchers, whose identities have not been publicly disclosed by their university, were running an experiment attempting to use large language models (LLMs) to sway users’ minds. The subreddit allows users to post opinions, and then rewards commenters with a delta (Δ) if their argument changes the original poster’s mind. In [one example](#), a bot reportedly took on the persona of a victim of sexual assault and argued that it “wasn’t that bad.”

The experiment [led](#) to community outrage, a formal investigation from the University of Zurich—which subsequently adopted a stricter ethics review—and the research itself was withheld from publication.

“Many people felt personally violated, and it inflamed existing tensions and anxieties about AI,” an anonymous moderator of the CMV subreddit shares, adding that most community members felt it was “a serious and appalling offense.”

Social media sites like X, Reddit, and other publicly available fora have been commonly mined for data by researchers even though [most users](#) are unaware. But how to run these studies ethically is still a challenging question that researchers grapple with.

Consent in Internet Research

Amy Bruckman, PhD, Regents’ professor and senior associate chair in the School of Interactive Computing at the Georgia Institute of Technology, says that the rules that

govern human research permit studying publicly available content.

“You could ask yourself, technically it’s allowed, but are you doing harm?” says Bruckman. “And the question of whether you’re doing harm really depends on what kind of content you’re looking at and how you’re using it.”

Over two decades ago, she authored research on the ethics of [disguising data](#) obtained from a public platform. With art, for example, users might want to be credited for their work, and anonymizing them without their consent is ethically wrong.

“

*You could ask yourself,
technically it’s allowed,
but are you doing harm?*

Amy Bruckman, PhD

A lack of explicit user consent, though, is common among research that uses Reddit posts. One [analysis](#) noted that, out of an original dataset of over 700 publications using Reddit data, only 134 studies even included discussion of the study’s ethics, such as a minimal statement that data were public or that an ethical review was unwarranted. The CMV moderator says that ethics boards and funders should advise against and decline experiments that do not consider “the most basic ethical principles” and obtain proper informed consent.

Additionally, while the United States has specific ethical frameworks and guidelines, they aren’t international.

“There are countries with absolutely no legislation protecting the use of human subjects in experimental studies,” says Bruckman. Because of the challenge involved in resolving these kinds of issues, she also formed the [ACM Special Interest Group on Computer-Human Interaction Research Ethics Committee](#), which confidentially advises researchers on these ethical questions.

Running an Ethical Internet Study

The moderator from the CMV subreddit says that it’s difficult to separate out any changes in engagement or tone on the subreddit after the experiment due to the general increase in LLM use.

“Regardless, there is a new element of suspicion at play where people now have to stop and question whether what they are reading was actually written by another person or not,” they write. “Suspicion breeds hostility and makes for a worse experience, so this is not ideal. Community members who were on the sub when the scandal occurred still allude to it sometimes.”

Part of the problem with the University of Zurich study is that it was poorly designed and hurting people, says Bruckman. “The benefits were approaching nothing,” she says. “So they were messing up people’s heads for no reason whatsoever.”

Bruckman, who is a moderator of /r/Science and other prominent subreddits, says that it isn’t mandatory to ask moderators on a public subreddit about scraping the data. But for “ephemeral” content that disappears, like Discord audio, it’s “wrong.” Anything that changes the experience for users, she adds, requires some consent.

Keywords: social media; ethics, research; research ethics; online social networking; internet; humans; data collection; biomedical research

There are examples of well-run studies that do this. Nathan Matias, PhD, a researcher at Cornell, worked with several /r/Science moderators to [test](#) whether posting the rules of a subreddit increased compliance. “Spoiler alert: it does,” says Bruckman. Because the study was well run, the results led to a change in Reddit corporate policy, which now has subreddits post their rules on the front page.

New Challenges

Although many people are unaware, researchers are mining public forums and social media for information. Conducting these studies unethically has in the past breached the trust of online communities. Though many researchers are careful about ethical considerations in their work, AI is adding an unexpected new challenge.

“Social media is reshaping the world, and here [conducting research on public forums] we have a window onto how,” says Bruckman. That window is now becoming harder for researchers to peek through.

In response to large AI companies scraping internet content, many companies like Reddit—whose corporate team have deemed the comments their intellectual property—have made it harder to access the data. “What you can and can’t do has been getting more and more restricted, so it’s actually harder than ever to datamine,” says Bruckman.

As public forums become less accessible, researchers will need to adopt new strategies to ethically engage with the individuals and information on these fora.

Please cite as:

Spichak S

How to Use—And How Not to Use—Public Forums in Research

J Med Internet Res 2026;28:e112727

URL: <https://www.jmir.org/2026/1/e112727>

doi: [10.2196/112727](https://doi.org/10.2196/112727)