

# The potentially adverse impact of Twitter 2.0 on scientific and research communication

December 8 2022, by Julia Cohen

---



Credit: Unsplash/CC0 Public Domain

In just over a month after the change in Twitter leadership, there have

been significant changes to the social media platform, in its new "Twitter 2.0." version.

For researchers who use Twitter as a primary source of data, including many of the computer scientists at USC's Information Sciences Institute (ISI), the effects could be debilitating.

## **Data for days with Twitter 1.0**

Over the years, Twitter has been extremely friendly to researchers, providing and maintaining a robust API (application programming interface) specifically for academic research. The Twitter API for Academic Research allows researchers with specific objectives who are affiliated with an academic institution to gather historical and real-time data sets of tweets, and related metadata, at no cost. Currently, the Twitter API for Academic Research continues to be functional and maintained in Twitter 2.0.

The data obtained from the API provides a means to observe public conversations and understand people's opinions about societal issues. Luca Luceri, a Postdoctoral Research Associate at ISI called Twitter "a primary platform to observe online discussion tied to political and [social issues](#)." And Twitter touts its API for Academic Research as a way for "academic researchers to use data from the public conversation to study topics as diverse as the conversation on Twitter itself."

However, if people continue deactivating their Twitter accounts, which appears to be the case, the makeup of the user base will change, with data sets and related studies proportionally affected. This is especially true if the user base evolves in a way that makes it more ideologically homogeneous and less diverse.

According to [MIT Technology Review](#), in the first week after its

transition, Twitter may have lost one million users, which translates to a 208% increase in lost accounts. And there's also the concern that the site could not work as effectively, because of the substantial decrease in the size of the engineering teams. This includes concerns about the durability of the service researchers rely on for data, namely the Twitter API. Jason Baumgartner, founder of Pushshift, a social media data collection, analysis, and archiving platform, said in several recent API requests, his team also saw a significant increase in error rates—in the 25-30% range –when they typically see rates near 1%. Though for now this is anecdotal, it leaves researchers wondering if they will be able to rely on Twitter data for future research.

One example of how the makeup of the less-regulated Twitter 2.0 user base could significantly be altered is if marginalized groups leave Twitter at a higher rate than the general user base, e.g. due to increased hate speech. Keith Burghardt, a Computer Scientist at ISI who studies hate speech online said, "It's not that an underregulated social media changes people's opinions, but it just makes people much more vocal. So you will probably see a lot more content that is hateful." In fact, a [study](#) by Montclair State University found that hate speech on Twitter skyrocketed in the week after the acquisition of Twitter.

## **The type of research at risk**

At USC's Information Sciences Institute, many scientists conduct research using data obtained from the Twitter API for Academic Research.

Katy Felkner, a graduate research assistant at ISI, studies artificial intelligence and language models. She used Twitter data sets to reduce anti-queer bias in AI by training a large language model using tweets written by members of the LGBTQ+ community. Additionally, she found that tweets from members of the LGBTQ+ community were

better at mitigating bias than tweets from outside that community about LGBTQ+ issues. She presented her resulting paper at the Queer in AI workshop at the North American Chapter of the Association for Computational Linguistics (NAACL) conference in July 2022.

Felkner explained why Twitter is so important to her work: "If you're getting data from the news, you're only getting the stories that are deemed newsworthy and a few perspectives on each story, whereas Twitter is very democratized and there is a low barrier of entry for a diverse set of participants. It's also very public, since most users have their tweets set to public. The Twitter API [for Academic Research] samples from all of the tweets on the platform at a certain time. So anyone who makes a tweet at time X about topic Y has some probability of being included in a data set about it."

Felkner pointed out, in addition to all of that, "it's kind of the last remaining text based social media platform." Facebook has text, but there's not a lot of public data; Instagram is photo-based; while TikTok is all videos. Felkner added, "extracting usable data from videos and images is often difficult and therefore prohibitively expensive in a research environment."

Kristina Lerman, a Principal Scientist at ISI, focuses on applying network and machine learning-based methods to problems in social computing. She currently has several projects that are using Twitter data. In one project, Lerman and her team are trying to identify social manipulation and influence campaigns on social media. She explained, "We are using Twitter data to see how malicious actors might be coordinating to affect public opinion in one way or another."

In other studies, she and Burghardt are using Twitter to identify factors that drive misinformation or anti-science attitudes. Lerman said, "We are collecting Twitter data to characterize the political ideology and how

much misinformation or anti-science content people are tweeting, to try to understand the roots of misinformation and discover who is susceptible to it." This complements work by Burghardt, who helped develop a method to predict anti-vaccine sentiment on Twitter, a problem that will very likely only get worse now that Twitter's vaccine misinformation policy is no longer enforced.

In yet another project, she is looking at gender identity and how people respond to and talk to people with different genders. Lerman says: "On Twitter, people do have some profile information; they can express their preferred pronouns. So unlike on other sites like Reddit, for example, where the profile information about the user identity is not revealed as much, we're relying on some functionality that's specific to Twitter about how people might express themselves and how much others might interact with them, based on their expression of their identity."

Given the changing nature of Twitter right now, Lerman and her team are in a bit of a precarious situation. She exclaimed, "We were just having discussions this morning about how we better hurry up and collect all of the data!" She gave an example, "In one project we are trying to understand how COVID authorities communicate. What kind of messaging strategies they use, and how people respond to that. So we're trying to hurry up and collect all the replies to the COVID authorities while we can."

Luceri is studying how misinformation spreads on Twitter and what can be done to prevent it. "A project we are currently working on is related to understanding how Twitter users are differently susceptible to misinformation, conspiracy theories, and online harms in general. In one of our recent papers we try to understand how people get radicalized to certain conspiracies, like QAnon."

The team wants both to detect deceptive and inauthentic activity, but

also to see how they can protect users from it. Luceri said, "We want to understand how Twitter users deal with fake news, misinformation, and conspiracy theory, and who the most vulnerable users are."

But they can't do that without the data. He explained, "The possibility that we will not have data, of course, is a problem, because our work leverages Twitter data sets and was also tailored for discovering things that might be helpful for Twitter itself." Luceri offered several specifics about the work he is doing, "We are looking to reveal the effectiveness of moderation policies, while observing users' engagement with harmful content. Our findings can inform social media providers, regulators, and policy makers to formulate strategies to counter the circulation of conspiracy theories and misinformation on social media. For example, understanding who are the most vulnerable users might allow Twitter to know how to deal with these users, and probably not expose them to all these attacks."

## **Impacts beyond data sets**

Jonathan May, ISI Research Team Leader, studies and teaches natural language processing (NLP), a subfield of AI concerned with how computers understand human language.

May has found Twitter to be professionally useful in ways beyond [data sets](#): "the international conversation about NLP has largely taken place on Twitter." He referenced one 2018 literal conversation that goes down in NLP Twitter history: the meaning/semantics mega-thread. Set into motion by Jacob Andreas, Assistant Professor at MIT, who tweeted about the ability of NLP models to understand meaning, it led to a whirlwind of academic debate and meaningful discussion in the NLP community. In fact, it was such a noteworthy thread, it has been written about and diagrammed. May said, "Twitter conversations tend to be open, and so the big open conversations take place there."

In the potential absence of Twitter-as-we-knew-it, May said discussions like these could find a new home. "There are a lot of essentially equivalent spaces. For example, Mastodon has a bit of a more decentralized nature to it." Several ISI researchers mentioned Mastodon as an academic Twitter alternative. The publication [Science reported](#) that many academics currently have their eyes on Mastodon, a free and decentralized social media platform that has a microblogging feature similar to Twitter.

May continued, "I think any sufficiently expressive social media space could do it. It's just kind of a matter of coming to a consensus that'll sort of evolve naturally based on—who knows?—whatever it was that allowed Twitter to become Twitter."

Provided by University of Southern California

Citation: The potentially adverse impact of Twitter 2.0 on scientific and research communication (2022, December 8) retrieved 26 April 2024 from <https://techxplore.com/news/2022-12-potentially-adverse-impact-twitter-scientific.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>
--