World view

By Ethan Zuckerman

Demand five precepts to aid social-media watchdogs

Facebook's heavy-handedness could open a window for research in the public interest.

Early this August, Facebook shut down the personal and organizational accounts of researchers associated with New York University's Ad Observatory, a project in which informed volunteers allow study of advertising targeted to their accounts. Facebook said its move was necessary to "protect people's privacy" and to comply with orders from the Federal Trade Commission. The FTC gave an unusually public response. It published a statement saying that its restrictions do not bar "good-faith research in the public interest".

This marks an opportunity for anyone who thinks that social media's effects on democracy and society should be open to scrutiny. It is time to lay down ground rules to empower public-interest research on social media.

In a collaboration with Elizabeth Hansen Shapiro at the Tow Center for Digital Journalism in New York City, I and other colleagues interviewed dozens of researchers, journalists and activists who study how social-media platforms affect democratic participation. Almost all named barriers to data access as a major obstacle, even those who helped to design Social Science One, a highly touted academia–industry partnership to study the spread of misinformation.

Researchers have techniques for dealing with the lack of information the platforms provide, although many such techniques are vulnerable to legal threats or restrictions. Ad Observatory asks for 'data donation' from a panel of web users who install a plug-in that allows researchers to study some aspects of the web users' online activity.

Another technique involves scraping — automated collection of content that appears to the general public or logged-in social-media users. This produces data sets such as PushShift, the most comprehensive archive of content available on the Reddit online discussion forum. Another is Media Cloud, a project I maintain with colleagues at several institutions to index millions of news stories a day and allow study of word frequencies over time. Its automated retrieval and data-storage features are technically identical to a search engine's, and thus prohibited by the non-negotiable terms of service required by most social-media platforms.

Until 2020, the United States' troublingly vague Computer Fraud and Abuse Act made researchers who violated a website's terms of service vulnerable to felony charges. That year, academic researchers argued successfully that using multiple social-media accounts to audit for discrimination should not be considered a criminal activity. A federal court agreed that "mere terms-of-service violations" do not merit criminal charges.

Although the ruling is welcome, uncertainty for

> **It is time to lay down ground rules to empower public-interest research of social media.**

Ethan Zuckerman is director of the Initiative for Digital Public Infrastructure at the University of Massachusetts, Amherst. e-mail: ethanz@ umass.edu

researchers remains, and social-media companies actively hinder their work. The FTC's endorsement of 'good-faith research' should be codified into principles guaranteeing researchers access to data under certain conditions.

I propose the following. First, give researchers access to the same targeting tools that platforms offer to advertisers and commercial partners. Second, for publicly viewable content, allow researchers to combine and share data sets by supplying keys to application programming interfaces. Third, explicitly allow users to donate data about their online behaviour for research, and make code used for such studies publicly reviewable for security flaws. Fourth, create safe-haven protections that recognize the public interest. Fifth, mandate regular audits of algorithms that moderate content and serve ads.

In the United States, the FTC could demand this access on behalf of consumers: it has broad powers to compel the release of data. In Europe, making such demands should be even more straightforward. The European Data Governance Act, proposed in November 2020, advances the concept of "data altruism" that allows users to donate their data, and the broader Digital Services Act includes a potential framework to implement protections for research in the public interest.

Technology companies argue that they must restrict data access because of the potential for harm, which also conveniently insulates them from criticism and scrutiny. They cite misuse of data, such as in the Cambridge Analytica scandal (which came to light in 2018 and prompted the FTC orders), in which an academic researcher took data from tens of millions of Facebook users collected through online 'personality tests' and gave it to a UK political consultancy that worked on behalf of Donald Trump and the Brexit campaign. Another example of abuse of data is the case of Clearview AI, which used scraping to produce a huge photographic database to allow federal and state law-enforcement agencies to identify individuals.

These incidents have led tech companies to design systems to prevent misuse — but such systems also prevent research necessary for oversight and scrutiny. To ensure that platforms act fairly and benefit society, there must be ways to protect user data and allow independent oversight.

Part of the solution is to create legal systems, not just technical ones, that distinguish between bad intent and legitimate, public-spirited research that can help to uncover social media's effects on economies and societies.

The influence of social-media companies is undeniable, and executives such as Facebook co-founder Mark Zuckerberg sincerely believe that their platforms make the world a better place. But they have been unwilling to give researchers the data to demonstrate whether this is so. It is time for society to demand access to those data.