Facebook made big mistake in data it provided to researchers, undermining academic work

Timberg, Craig

🔗 ProQuest document link

FULL TEXT

Facebook provided a data set to a consortium of social scientists last year that had serious errors, affecting the findings in an unknown number of academic papers, the company acknowledged Friday.

The company used a regular monthly call on Friday with roughly three dozen researchers affiliated with Social Science One, a consortium founded in 2018 that Facebook hails as a model for collaboration with academics, to admit the error and apologize for the impact on their work.

The data concerns the effect of social media on elections and democracy and includes what web addresses Facebook users click on, along with other information.

The error resulted from Facebook accidentally excluding data from U.S. users who had no detectable political leanings —a group that amounted to roughly half of all of Facebook's users in the United States. Data from users in other countries was not affected.

"It's data. Of course, there are errors," said Gary King, a Harvard professor who co-chairs Social Science One. "This, of course, was a big error."

King, director of the university's Institute for Quantitative Social Science, said dozens of papers from researchers affiliated with Social Science One had relied on the data since Facebook shared the flawed set in February 2020, but he said the impact could be determined only after Facebook provided corrected data that could be reanalyzed. He said some of the errors may cause little or no problems, but others could be serious.

Social Science One shared the flawed data with at least 110 researchers, King said.

The group's former co-chairman, Stanford Law professor Nathaniel Persily, said of the incident: "This is a friggin' outrage and a fundamental breach of promises Facebook made to the research community. It also demonstrates why we need government regulation to force social media companies to develop secure data sharing programs with outside independent researchers."

An Italian researcher, Fabio Giglietto, discovered data anomalies last month and brought them to Facebook's attention. The company contacted researchers in recent days with news that they had failed to include roughly half of its U.S. users —a group that likely is less politically polarized than Facebook's overall user base. The New York Times first reported Facebook's error.

"This issue was caused by a technical error in our URL Shares Data Set, which we proactively told impacted partners about and are working swiftly to resolve," Facebook spokeswoman Mavis Jones said.

The anonymized data set is one of the largest in social science history, with 42 trillion numbers. The set includes protections against individual users being identified based on what they have posted on Facebook, King said. He said the company began working more closely with researchers after the Cambridge Analytica scandal in 2018, but there have been tensions with researchers over how much information is shared by the company, which often cites privacy concerns when not providing data with the granularity they desire for their work.

Cody Buntain, a member of the consortium and an assistant professor of informatics for the New Jersey Institute of Technology, said richer data from Facebook would have allowed researchers to discover the error sooner through routine checks. He said he was directly aware of several papers whose data now would need reanalyzing. There's no immediate timetable for Facebook to provide the corrected data, which is so large that it typically takes weeks to

process.

"This is a totally foreseeable, preventable problem," Buntain said.

DETAILS

Subject:	Researchers; Consortia; Datasets; Science; Social sciences; Social networks
Business indexing term:	Subject: Social networks
Location:	United States--US
Publication title:	The Washington Post (Online); Washington, D.C.
Publication year:	2021
Publication date:	Sep 10, 2021
Section:	Technology
Publisher:	WP Company LLC d/b/a The Washington Post
Place of publication:	Washington, D.C.
Country of publication:	United States, Washington, D.C.
Publication subject:	General Interest Periodicals--United States
ISSN:	26419599
Source type:	Blog, Podcast, or Website
Language of publication:	English
Document type:	News
ProQuest document ID:	2571349703
Document URL:	https://pitt.idm.oclc.org/login?url=https://www.proquest.com/blogs-podcasts-websites/facebook-made-big-mistake-data-provided/docview/2571349703/se-2?accountid=14709
Copyright:	Copyright WP Company LLC d/b/a The Washington Post Sep 10, 2021
Last updated:	2021-09-11
Database:	U.S. Major Dailies

LINKS

CHECK ARTICLE AVAILABILITY