MIT Technology Review

MIT Technology Review

Artificial intelligence / Machine learning

How to poison the data that Big Tech uses to surveil you

Algorithms are meaningless without good data. The public can exploit that to demand change.

by **Karen Hao**

March 5, 2021

ERIC RISBERG / AP

Every day, your life leaves a trail of digital breadcrumbs that tech giants use to track you. You send

target you with ads and recommendations. Google cashes your data in for over $120 billion a year of ad revenue.

Increasingly, we can no longer opt out of this arrangement. In 2019 Kashmir Hill, then a reporter for Gizmodo, famously tried to cut five major tech giants out of her life. She spent six weeks being miserable, struggling to perform basic digital functions. The tech giants, meanwhile, didn't even feel an itch.

Now researchers at Northwestern University are suggesting new ways to redress this power imbalance by treating our *collective* data as a bargaining chip. Tech giants may have fancy algorithms at their disposal, but they are meaningless without enough of the right data to train on.

In a new paper being presented at the Association for Computing Machinery's Fairness, Accountability, and Transparency conference next week, researchers including PhD students Nicholas Vincent and Hanlin Li propose three ways the public can exploit this to their advantage:

- **Data strikes**, inspired by the idea of labor strikes, which involve withholding or deleting your data so a tech firm cannot use it—leaving a platform or installing privacy tools, for instance.
- **Data poisoning**, which involves contributing meaningless or harmful data. AdNauseam, for example, is a browser extension that clicks on every single ad served to you, thus confusing Google's ad-targeting algorithms.
- **Conscious data contribution**, which involves giving meaning*ful* data to the competitor of a platform you want to protest, such as by uploading your Facebook photos to Tumblr instead.

People already use many of these tactics to protect their own privacy. If you've ever used an ad

But as Hill found, sporadic individual actions like these don't do much to get tech giants to change *their* behaviors.

What if millions of people were to coordinate to poison a tech giant's data well, though? That might just give them some leverage to assert their demands.

There may have already been a few examples of this. In January, millions of users deleted their WhatsApp accounts and moved to competitors like Signal and Telegram after Facebook announced that it would begin sharing WhatsApp data with the rest of the company. The exodus caused Facebook to delay its policy changes.

Just this week, Google also announced that it would stop tracking individuals across the web and targeting ads at them. While it's unclear whether this is a real change or just a rebranding, says Vincent, it's possible that the increased use of tools like AdNauseam contributed to that decision by degrading the effectiveness of the company's algorithms. (Of course, it's ultimately hard to tell. "The only person who really knows how effectively a data leverage movement impacted a system is the tech company," he says.)

Vincent and Li think these campaigns can complement strategies such as policy advocacy and worker organizing in the movement to resist Big Tech.

"It's exciting to see this kind of work," says Ali Alkhatib, a research fellow at the University of San Francisco's Center for Applied Data Ethics, who was not involved in the research. "It was really interesting to see them thinking about the collective or holistic view: we can mess with the well and make demands with that threat, because it is our data and it all goes into this well together."

There is still work to be done to make these campaigns more widespread. Computer scientists could play an important role in making more tools like AdNauseam, for example, which would help lower the barrier to participating in such tactics. Policymakers could help too. Data strikes are most effective when bolstered by strong data privacy laws, such as the European Union's General Data Protection Regulation (GDPR), which gives consumers the right to request the deletion of their data. Without such regulation, it's harder to guarantee that a tech company will give you the option to scrub your digital records, even if you remove your account.

Related Story

This is how we lost control of our faces

The largest ever study of facial-recognition data shows how much the rise of deep learning has algorithm? And what kind of data would be most effective in poisoning a particular system? In a simulation involving a movie recommendation algorithm, for example, the researchers found that if 30% of users went on strike, it could cut the system's accuracy by 50%. But every machine-learning system is different, and companies constantly update them. The researchers hope that more people in the machine-learning community can run similar simulations of different companies' systems and identify their vulnerabilities.

Alkhatib suggests that scholars should do more research on how to inspire collective data action as well. "Collective action is really hard," he says. "Getting people to follow through on ongoing action is one challenge. And then there's the challenge of how do you keep a group of people who are very transient—in this case it might be people who are using a search engine for five seconds—to see themselves as part of a community that actually has longevity?"

These tactics might also have downstream consequences that need careful examination, he adds. Could data poisoning end up just adding more work for content moderators and other people tasked with cleaning and labeling the companies' training data?

But overall, Vincent, Li, and Alkhatib are optimistic that data leverage could turn into a persuasive tool to shape how tech giants treat our data and our privacy. "AI systems are dependent on data. It's just a fact about how they work," Vincent says. "Ultimately, that is a way the public can gain power."

Share Link

Author Karen Hao

Biotechnology Mar 18

Did the coronavirus leak from a lab? These scientists say we shouldn't