

As Facebook embraces artificial intelligence tools, will it further spook consumers?

April 17 2018, by Tim Johnson, McClatchy Washington Bureau

Social media companies have embraced artificial intelligence tools to scrub their platforms of hate speech, terrorist propaganda and other content deemed noxious. But will those tools censor other content? Can a program judge the value of speech?

Facebook founder Mark Zuckerberg told Congress last week that his [company](#) is rapidly developing AI tools to "identify certain classes of bad activity proactively and flag it for our team."

It is one of several moves by Facebook as it struggles with an erosion of consumer trust over its harvesting of user data, its past vulnerability to targeted political misinformation and the opaqueness of the formulas upon which its news feeds are built.

Some technologists believe that AI tools won't resolve these issues that Facebook and other social media companies face.

"The problem is that surveillance is Facebook's business model: surveillance in order to facilitate psychological manipulation," said Bruce Schneier, a well-known security expert and privacy specialist. "Whether it's done by people or ([artificial intelligence](#)) is in the noise."

Zuckerberg said his Menlo Park, Calif., company relies on both AI tools and thousands of employees to review content. By the end of the year, he said, some 20,000 Facebook employees will be "working on security and content review."

The company is developing its AI tools to track down hate speech and fake news on its platform and views the tools as a "scalable way to identify and root out most of this harmful content," he said, indicating several times over 10 hours of testimony on two days that Facebook algorithms can find objectionable content faster than humans.

"Today, as we sit here, 99 percent of the ISIS and al-Qaida content that we take down on Facebook, our AI systems flag before any human sees it," Zuckerberg said at a Senate hearing, referring to extremist Islamic groups.

The [artificial intelligence systems](#) work in conjunction with a counterterrorism team of humans that Zuckerberg says numbers 200 employees. "I think we have capacity in 30 languages that we're working on," he said.

Other existing AI tools "do a better job of identifying [fake accounts](#) that may be trying to interfere in elections or spread misinformation," he said. After fake accounts placed political information on Facebook that disrupted the 2016 election, Facebook proactively took down "tens of thousands of fake accounts" before French and German elections in 2017, and Alabama's special election for a vacant Senate seat last December, he added.

Facebook is far from alone among social media companies harnessing artificial intelligence to assist humans monitoring content.

"AI tools in concert with humans can do better than either can do alone," said Wendell Wallach, an investigator at The Hastings Center, a bioethics research institute in Garrison, New York.

But Wallach noted that many users do not understand artificial intelligence, and Big Tech may face a backlash like food concerns face

over genetically modified (GMO) ingredients.

"The leading AI companies, which happen to be the same as the leading digital companies at the moment, understand that there is a GMO-like elephant that could jump out of the AI closet," Wallach said.

Already, concern is mounting among conservatives on Capitol Hill that platforms like Facebook tilt to the political left, whether AI tools or humans are involved in making content decisions.

"You recognize these folks?" Rep. Billy Long, R-Mo., asked Zuckerberg while holding up a photo of two sisters.

"Is that Diamond and Silk?" Zuckerberg asked, referring to two black social media personalities who are fervent supporters of President Donald Trump.

Indeed, it was, Long said, and Facebook had deemed them "unsafe."

"What is unsafe about two black women supporting President Donald J. Trump?"

Zuckerberg later noted that his Facebook team "made an enforcement error, and we've already gotten in touch with them to reverse it."

Artificial intelligence tools excel at identifying salient information out of masses of data but struggle to understand context, especially in spoken language, experts said.

"The exact same sentence, depending on the relationship between two individuals, could be an expression of hate or an expression of endearment," said David Danks, an expert on ethics around autonomous systems at Carnegie Mellon University. He cited the use of the "N-

word," which between some people can be a friendly term, but is also widely considered [hate speech](#) in other contexts.

Any errors that AI tools make in such linguistic minefields could be interpreted as censorship or political bias that could further diminish trust in social media companies.

"The general public, I think, is much less trusting of these companies," Danks said.

Eventually, he said, the algorithms and AI tools of a handful of companies will earn greater public trust, even as consumers do not understand how they operate.

"I don't understand in many ways how my car works but I still trust it to function in all the ways I need it to," Danks said.

Just as librarians once used subjective judgment in taking books off the shelves, social media companies also can face criticism that AI tools can overreach.

"Twitter faces this," said James J. Hughes, executive director of the Institute for Ethics and Emerging Technologies, in Boston. "Pinterest and Instagram are always taking down artists' websites who happen to have naked bodies in them when they think they are porn, when they are not.

"And they are doing that based on artificial intelligence algorithms that flag how much naked flesh is in the picture."

In his testimony, Zuckerberg said AI tools were increasingly adept at "identifying fake accounts that may be trying to interfere in elections or spread misinformation."

Facebook has admitted that a Russian agency used Facebook to spread misinformation that reached up to 126 million people around the time of the 2016 presidential vote, and that the personal data of 87 million people may have been misused by the firm Cambridge Analytica to target voters in favor of Trump.

Zuckerberg told senators that Facebook's delay in identifying Russian efforts to interfere in the election is "one of my greatest regrets in running the company" and pledged to do better to combat manipulation for this year's election.

As legislators wrestled over whether Facebook and other social media companies need regulation, Zuckerberg repeatedly faced questions over the nature of his company. Is it also a media company because it produces content? A software company? A financial services firm that supports money transfers?

"I consider us to be a technology company, because the primary thing that we do is have engineers who write code and build products and services for other people," Zuckerberg told a House hearing.

Experts say that answer doesn't address complex issues around platforms that are increasingly more like public utilities.

"The electric company is not allowed to say, we don't like your political views therefore we are not going to give you electricity," Danks said. "If somebody is knocked off of Facebook, is that tantamount to the electric company cutting off their electricity? Or is it more like the person who is really loud and obnoxious in a bar, and the owner says, 'You need to leave now'?"

©2018 McClatchy Washington Bureau
Distributed by Tribune Content Agency, LLC.

Citation: As Facebook embraces artificial intelligence tools, will it further spook consumers? (2018, April 17) retrieved 3 May 2024 from <https://phys.org/news/2018-04-facebook-embraces-artificial-intelligence-tools.html>

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.