The Ethics of AI for Suicide Prevention

Ariella Brown, Technology Writer
12/13/2017 00:01 AM EST

Facebook is using artificial intelligence to signal deviations from normal posting behavior to provide warning of possible suicidal tendencies. Is that okay?

In today's hyperconnected world, we are generating and collecting so much data that it is beyond human capability to sift through it all. Indeed, one application of artificial intelligence is identifying patterns and deviations that signal intent on posts. Facebook is using AI in this way to extract value from its own Big Data trove. While that may be applied to a good purpose, it also raises ethical concerns.

Where might one get insight into this issue? In my own search, I found an organization called PERVADE (Pervasive Data Ethics for Computational Research). With the cooperation of six universities and the funding it received this September, it is working to frame the questions and move toward the answers.

I reached out to the organization for some expert views on the ethical questions related to Facebook's announcement that it was incorporating AI in its expanded suicide-signal detection effort. That led to a call with one of the group's members, Matthew Bietz.

Bietz told me the people involved in PERVADE are researching the ramifications of pervasive data, which encompasses continuous data collection — not just from what we post to social media, but also from the "digital traces that we leave behind anytime we're online," such as when we Google or email. New connections from the Internet of Things (IoT) and wearables further contribute to the growing body of "data about spaces we're in," he said. As this phenomenon is "relatively new," it opens up new questions to explore with respect to "data ethics."

Data detection and the EU
Noting that Facebook had declared it would not extend its suicide-signal-detection tools into the European Union, I asked Bietz whether the program would fall afoul of the EU's General Data Protection Regulation. He acknowledged at the EU has "some of the strictest data regulations in the world" but added that "it is not entirely clear that what Facebook is doing" would be illegal under the GDPR standard.

It might be more accurate to say that Facebook is venturing onto "an edge that hasn't really been tested," he said, "and my guess is they decided they don't want to be the test case."

Facebook's terms of service give it permission to look at data, Bietz said, so that activity doesn't automatically violate the GDPR privacy regulations. But the legislation prohibits companies from profiling based solely "on algorithms or just automatically by a computer."

That means that if the algorithm makes the call on a signal "without people being involved" in determining what the signal might mean, such activity would be on the wrong side of the GDPR. It is possible, though, to stay within the legal limits by using the algorithm only "to help decide which posts need human eyes" on them, he said.

Aside from questions of law, Bietz believes Facebook's approach to suicide detection sets up an "ethical minefield because the consequences are so big." That doesn't mean that what the company doing is "necessarily unethical," it really depends on "how they're handling the data" and whether Facebook "deserves our trust."

It's an open question, he said, whether Facebook can be relied on to use the suicide-signal data expressly for the purpose of preventing suicide and ameliorating that societal problem, when its profit model is based on selective information sharing. To prove its good intentions, Facebook should be forthcoming on "how far the information is being shared," exactly what they do with it, and how they keep it secure. But Facebook has taken the position that we just have to trust them on this.

Part of the problem is that we are now facing "new technologies and new applications of technologies at a scale that we haven't seen before," said Bietz. Consequently, there is no clear precedent to guide ethical decisions. We are "exploring new ground, not just technically, but ethically as well." And it takes some orientation just to frame "the right ethical questions to ask."

Questions and ramifications
Bietz said he is in favor of using data and algorithms to prevent suicide, pointing out that Crisis Text Line applied the method to prioritize responses to the most urgent cases (see Crisis Text Line: Saving Lives Through Data). Facebook has not clarified how it plans to prevent suicides, however, and Bietz thinks the company should not attempt this on its own. "I'd feel a lot better about it if I knew they were working closely with psychological researchers from outside the company who were thinking about some of these issues and working with them," he said.

The quality of the algorithms used is a critical issue, Bietz added: Erring on the side of excess caution, for example, could yield false positives. But Facebook has not communicated the nature of its algorithms or what it plans to do with them down the line. Will it stick with the original algorithms or tweak them over time? As it normally does, Facebook is keeping the details to itself.

Can we trust Facebook to be ethical? We have to ask that question, in light of the fact that the social media platform has access to so much information about us. Not only can it recognize our standard "pattern" of behavior from our history, but it can also infer information that we might not have shared, on the basis of friend posts.

Never before has society been subject to this kind of surveillance, which "allows pinpoint precision down to the level of the individual" combined with the huge range of "the multiple millions who have Facebook accounts," Bietz observed. It's a brave new world of data that compels us to work on finding "where the ethical lines are."

And that's what PERVADE — in cooperation with a number of other organizations — intends to work on.