

January 15, 2019

Medscape

Facebook's AI Suicide Prevention Program: Likes and Dislikes

By Deborah Brauser

Within the first year of launching an artificial intelligence (AI) program to flag potential signs of suicidality in individual accounts, Facebook has released data showing there have been 3500 wellness checks worldwide to rule out potential suicidal intent in its users.

Although at first blush this sounds like a good thing, some experts are questioning specifics of the algorithm and calling on the company for greater transparency.

Initiated in 2017, the program uses machine learning to identify posts, comments, and videos that may indicate suicidal intent. It also uses "contextual understanding" to disregard phrases such as, "I have so much homework I want to die," which is not a genuine sign of distress, the company said in a press release.

Highly concerning content is then reviewed by specialist teams who determine whether specific individuals warrant help from first responders.

"Mental health experts say that one of the best ways to prevent suicide is for people in distress to hear from friends and family who care about them," a spokesperson for Facebook said in a statement sent to Medscape Medical News. "We can connect those in distress with friends (and also organizations) who can offer support."

Asked to comment, Drew Ramsey, MD, assistant clinical professor of psychiatry at Columbia University, New York City, and member and past chair of the American Psychiatric Association's (APA's) Council on Communications, said the company should be commended for working toward curbing growing suicide rates.

"I think Facebook is stepping up out of a responsibility that their platform is being used by people to express a lot of different feelings. And when that happens, they wanted to do their best to respond," Ramsey told Medscape Medical News.

However, the company has not released specific data from its program, including how many of the reports turned out to be actual emergency situations.

In an op-ed published in the Washington Post in December, Mason Marks, MD, JD, visiting fellow at Yale Law School, New Haven, Connecticut, said that, although revolutionary, this type of suicide prevention technology "badly needs oversight."

Ethical Issues?

Former APA President Paul S. Appelbaum, MD, director of the Division of Law, Ethics, and Psychiatry at Columbia University, New York City, has even greater concerns.

"In principle, using advanced AI or machine learning techniques to identify people at increased risk for suicidality is not

inherently problematic in its own right, to the extent that it might in fact allow us to identify people who wouldn't otherwise be considered to be at risk and to allow interventions with them," he told Medscape Medical News.

"In principle, the effort is worthwhile and perhaps even commendable. But what's of concern is that it's being undertaken by a for-profit entity in a 'covert' way—and I don't think that word is too strong to use.

"The algorithm used is proprietary; its accuracy and how accuracy has been established are unknown; the use that Facebook will make of that information is unclear; and their track record in protecting their users' information from third parties is not encouraging," added Appelbaum, who is also a former chair of the APA's Council on Psychiatry and Law and a current co-chair of the Standing Committee on Ethics for the World Psychiatric Association.

He noted that questions remain about the process of identifying and possibly helping at-risk individuals.

"All of these things are reasons for considerable concern. If this were to be done in a socially responsible way, it would be an open process," he said.

As such, the algorithm would be publicly available and subject to critique, Facebook would reach out to leading suicide researchers and involve them in their process, an effort would be made to openly validate the algorithm being used, and an open strategy would be developed for responding to those with indicators of elevated risk, said Appelbaum.

"But so far, it does not sound as though Facebook is doing this the way it should be done," he noted.

"Black Boxed" Process?

John Torous, MD, director of the Division of Digital Psychiatry at Beth Israel Deaconess Medical Center, Boston, Massachusetts, noted that any effort to decrease suicide is a good step but agreed with Appelbaum that this particular effort needs more transparency.

"There is certainly a need for innovative approaches towards suicide prevention; and it's exciting to see potential technology to help reduce suicide. But it's hard to comment on what this program is because there's so little that's actually known," Torous told Medscape Medical News.

He added that the specific data that go into the algorithm and what the algorithm is actually doing are, at this point, unknown.

"Also, just because you send an ambulance to someone's house, was that necessary? Was that right? Did it help? In essence, we have very little data, so it's challenging to even evaluate it. But it's a public health problem, and you want a lot of transparency on this," Torous said.

"You want to shine the light of transparency so people understand and trust this. You have to ask, why is this so black boxed when it's such a public issue? I think they want to do the right thing and mean well, but their data and [posted] videos don't really explain anything," he said.

Torous, who leads a work group on the evaluation of smartphone apps for the APA and has a background in computer science and engineering, noted that it's important for clinicians to stay up to date on technology and social media.

"Clinicians need to become increasingly aware of these different digital health initiatives, whether good or bad. These are things that can't be ignored because they're going to impact patients and the practice of medicine," he said.

"Even if you're a clinician that doesn't want to use social media or you use it and don't like it, you still need to be

informed. It's also important that clinicians begin shaping policies about this," Torous added.

"You have to ask, why is Facebook taking the lead in issues like these? It's important for clinicians to raise the important questions, pose new solutions, and work on new solutions ourselves."

"Devil in the Details"

Also commenting for Medscape Medical News, [Ipsit Vahia, MD](#), medical director of the [McLean Institute for Technology in Psychiatry](#), Belmont, Massachusetts, said suicide prevention efforts are greatly needed.

"Suicide has been trickling up, and, combined with substance use, the American lifespan actually decreased this year for the first time, as per the Centers for Disease Control and Prevention [CDC]," said Vahia, who is also medical director of geriatric psychiatry outpatient services at McLean Hospital.

"So I think anything we can do leveraging technology to impact [the suicide rate] is welcome. However, the devil is in the details. There is not enough evidence to tell us exactly how successful this is and what the preventive rate is. The danger in an approach like this is that there are too many false positives—and that can impact our trust in the technology itself," he said.

In 2017, there were more than 47,000 deaths by suicide and more than 1.3 million suicide attempts in the United States, according to the CDC. In addition, the suicide rate increased dramatically by 33% from 1999 to 2017.

Some social media users have used live streaming services, such as Facebook Live, to broadcast their suicides. According to the New York Post, this very thing was done by a 14-year-old girl in Florida in January 2018.

The company's AI program scans user content and comments for possible suicide risk, but Facebook users themselves can also report on posts or videos (including livestreaming) in which someone else appears to be in distress.

After a report is made or the AI program flags something, it is reviewed by the company's community operations team. This team consists of "thousands of people around the world," all of whom have training on both suicide and self-harm.

"This piecing together of an international network struck me as innovative. It's a way to rapidly deploy services all around the world," Ramsey said.

"Only content that is highly correlated to immediate self-harm or imminent risk of suicide is [then] reviewed by our specialist teams," the company explained. These teams are trained to work directly with first responders and have backgrounds in such fields as US law enforcement and suicide hotlines.

Surveillance State?

In its statement to Medscape Medical News, Facebook noted that instances in which it believes there is immediate risk for serious harm resulting in a welfare check by first responders represent a small minority of the flagged cases.

The company added that escalating "everything" to first responders could result in all escalations not being taken seriously.

Although it said that it helped first responders reach about 3500 people globally for wellness checks during the AI's first year of use, Facebook did not respond to questions by Medscape Medical News about whether it will be releasing more details, including how many of the reports were actual emergencies and whether it is expecting to make changes to the program as it heads into its second year.

After going through a false positive experience, “someone could think the mental health system is a surveillance state, where if you express something in what you think is an anonymous way, you could get the police knocking on your door. And that’s a pretty frightening prospect,” Torous noted.

In a story published by NPR in November, Antigone Davis, Facebook global head of safety, said that releasing too many details about the AI program could let individuals “play games with the system.”

However, “if it’s a robust system, it shouldn’t be that fragile,” Torous said. “Are they saying the system is that unreliable that slight changes will perturb it and change the predictions? If you’re sending out an ambulance, you want to be basing it on some pretty good evidence.”

Ramsey countered that erring on the side of caution isn’t necessarily a bad thing.

“As a psychiatrist and a public mental health advocate, I make false positive wellness calls all the time. In terms of the oversight [of Facebook’s AI] and how that should be handled, that’s a good question,” he added.

“It raises concerns about ‘big brother’; but when big brother is calling 911 because individuals on Facebook Live are self-harming or considering suicide, it’s definitely a new chapter in [this] epidemic and how we’re responding.”

Ramsey also noted that someone who is worried about privacy concerns probably wouldn’t be posting on a public social media site.

Stigma, Involuntary Interventions

On the other hand, Appelbaum said that although “efforts are badly needed” to address the dramatic spike in suicide in the United States, “it needs to be addressed in a thoughtful and open manner.

“Efforts to prevent suicide could, if not done right, carry with them the risk of stigma or involuntary interventions in the lives of people who may not warrant such interventions. So there are significant consequences for both error and misuse of the information that is gathered,” he said.

Another concern is that fear of monitoring may make some users back away from reaching out to friends on social media during times of crisis, Appelbaum said.

“People may not want a Facebook record of suicidality or to be labeled in that way. So they may choose not to use it to reach out to what might be their natural support system, leaving them worse off,” he said.

Torous noted that some people also adopt social media-specific personas. “There’s knowledge that what people post on Facebook is often a happier picture, such as vacation photos and showing themselves always having a good time. So it’s hard to know if we’d be responding to a person’s true mental state or if it’s their online state,” he said.

“It opens up questions as to what should we really be responding to, and what are the consequences,” he said.

Torous added that there’s also a question about where young people are actually congregating online. “Where are they truly expressing how they feel and sharing genuine emotion that’s important to respond to? It shifts so rapidly.”

Facebook owns Instagram, but Snapchat and instant messaging apps such as Kik, as well as online video services such as YouTube, are also popular with young people.

“I have to wonder: is this Facebook algorithm even reaching the people who most need help? I don’t think we know,” Torous said.

New Tools

That said, “there is a significant interest in the academic psychiatric community around how we can use these new tools to help save lives,” said Ramsey.

He noted that “lots of other companies” are using technology to try to prevent self-harming behaviors. “There are many who are using the tremendous amount of data that our phones generate to predict all sorts of behavior,” he said.

For example, a prolific texter who stops texting friends except for short, negative messages, starts playing sad songs, and doesn’t leave the house for a long period may be flagged regarding concerning behaviors, he added.

“As a clinician, that’s the kind of data that of course I would want to know about my patients. It’s definitely a new set of tools; and like any new tool, it’s going to take us a little while to understand how to most effectively use them,” he said.

Vahia reported that “there are several efforts in addition to Facebook’s” to synthesize data through the use of natural-language processing to pick up on signals of declining mood and suicide.

However, “there are several questions around privacy and what extent using AI on public posts constitutes violating privacy and whether it’s done with people’s consent,” he said.

“I think if this is done without properly addressing this, it’s going to raise questions as to whether people can trust AI technology and whether privacy is invaded if signals are being picked up on,” Vahia said.

In addition, “any technology or medication in healthcare needs to have a very high degree of sensitivity and specificity. The higher the degree, the more likely they’re going to be trusted by clinicians and patients,” he said. “This technology has high potential, but I think it needs refining.”

“As far as Facebook wanting to reduce rates of suicide, that’s definitely commendable. However, to date, they have not gone about it in the right way and have created more concerns—both about the effectiveness of their approach and about the use of the information that they are gathering,” said Appelbaum.

“The right way to do this would be in an open fashion, collaborating with the community of suicide researchers and creating algorithms that could be shared more widely so other social media entities and suicide researchers could have access to them. They could do an enormous public service here that they have so far chosen not to undertake,” he added.

Torous said if Facebook wants to “kind of take on this public health role,” it should also work on securing individuals’ data.

“Where is Facebook sending this information to? Who has access to it? Where and how are they storing it? What if that’s attacked and there’s a lift of people that Facebook thinks is suicidal? Transparency is the answer here, along with consent. Having neither brings up a host of ethical issues,” he said.

Overall, Appelbaum agreed that staying up to date on technologic advances and social media news is important for clinicians.

“The development of algorithms such as Facebook’s means that nonclinical entities may have equal or even greater access to the mental states of patients. So it is important for clinicians to have input into how that information is gathered and what is done with it. They know best what the risks are and how to maximize the benefits,” he said.

“Clearly, for better or worse, this is happening. And it’s time for the clinical community to make sure they’re actively engaged stakeholders in all of this,” Torous concluded.