

Is A Threat On Facebook Real? Supreme Court Will Weigh In

by Aarti Shahani

This week, the Supreme Court agreed to hear the case of a man on Facebook who threatened to kill his wife.

In 2010, Pennsylvania resident Anthony Elonis got dumped, lost his job and expressed his frustrations via the Internet.

"He took to Facebook as a form of, what he says, a form of therapy," says criminologist Rob D'Ovidio of Drexel University, who is following the case.

Is It A 'True Threat'?

As life kept falling apart, Elonis made threats repeatedly on Facebook to his ex, to law enforcement and to an unspecified elementary school. He was convicted and sentenced to 44 months in prison and three years of supervised community release. D'Ovidio says that's a serious felony sentence.

Elonis' wife said she felt scared, but his defense says the graphic language was a joke. Take this post:

Did you know that it's illegal for me to say I want to kill my wife?

It's illegal.

It's indirect criminal contempt.

It's one of the only sentences that I'm not allowed to say.

Now it was okay for me to say it right then because I was just telling you that it's illegal for me to say I want to kill my wife.

Elonis claims he lifted the lines, almost word-for-word, from the show *The Whitest Kids U' Know*, in which comedian Trevor Moore begins "Did you know that it's illegal to say, 'I want to kill the president of the United States of America.' "

The Supreme Court will consider if Elonis' language was a "true threat," which the as speech that is so clearly objectionable, any objective listener could be scared.

Facebook: Context Matters

Meanwhile, Facebook has already decided that key words are not an effective way to look for threats on the site.

"Things that get reported for the more intense reasons are things that you look at the text and it's like, 'I had no idea from looking at this text that this was going on,'" says Arturo Bejar, a director of engineering at Facebook.

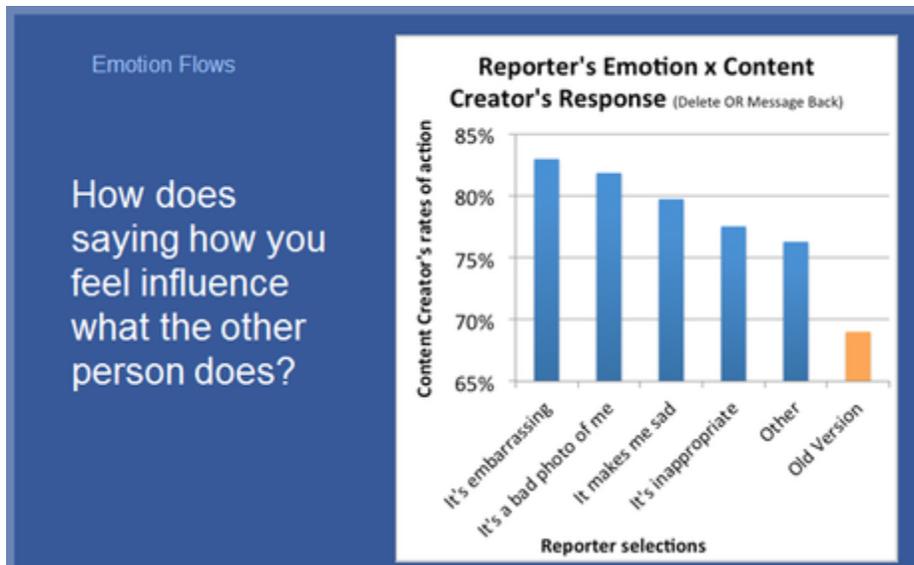
While the platform has against porn, it does not forbid specific violent words. While algorithms crawl through the site in search of our deepest consumer demands, there's no algorithm looking for credible threats. That's because "intent and perception really matter," Bejar says.

Bejar's little-known section of the Facebook machine works on . He has gone to leading universities and recruited experts in linguistics and compassion research.



Facebook is studying users' negative reactions to posts by 'friends.'

Facebook



In an effort to manage online conflict, Facebook is testing out words that prompt users to behave more compassionately with each other.

Facebook

Together, they field user complaints about posts at a massive scale. They facilitate "approximately 4 million conversations a week," Bejar says.

By conversation, he really does mean getting people to communicate directly with each other and not just complain anonymously. It's couples therapy-lite for the social media age. And, it turns out, a button that says "report" is a real conversation killer.

"We were talking to teenagers, and it turns out they didn't like clicking on 'report' because they were worried they'd get a friend in trouble," he says.

When his team changed it to softer phrases like "this post is a problem," complaints shot up.

They also revamped the automated form, so that the person complaining names the recipient and the emotion that got triggered. Let's say I hurt a friend's feelings. He could send a form letter: "Hey Aarti, this photo that you shared with me is embarrassing to me."

More people walked through the process of complaining. And according to the data, the word "embarrassing" really works.

"There's an 83 to 85 percent likelihood that the person receiving the message is going to reply back or take down the photo," Bejar says.

A Work In Progress

Facebook has hundreds of employees around the world who can step in when the automated tools fail, and threat detection is clearly a work in progress.

Consider two cases. In the first, Facebook user Sarah Lebsack complained about a picture that a friend posted of his naked butt.

"It wasn't the most attractive rear end I've ever seen, but also just not what I wanted to see as I browsed Facebook," Lebsack says. She says it took Facebook a couple of hours to take down the picture.

User Francesca Sam-Sin says she complained about a post that put her safety at risk. Recently she had flowers delivered to her mom after a surgery, and her mom posted a picture of the flowers.

"The card had my full name, my address and my cellphone number on it. And it was open to the public; it wasn't just limited to her friends," Sam-Sin says.

Sam-Sin says her mom wouldn't delete the post because she wanted to show off the bouquet, and Facebook wouldn't get involved in family matters.