The New York Times | https://nyti.ms/22ItcuN

TECHNOLOGY

Microsoft Created a Twitter Bot to Learn From Users. It Quickly Became a Racist Jerk.

By DANIEL VICTOR MARCH 24, 2016

Microsoft set out to learn about “conversational understanding” by creating a bot designed to have automated discussions with Twitter users, mimicking the language they use.

What could go wrong?

If you guessed, “It will probably become really racist,” you’ve clearly spent time on the Internet. Less than 24 hours after the bot, @TayandYou, went online Wednesday, Microsoft halted posting from the account and deleted several of its most obscene statements.

The bot, developed by Microsoft’s technology and research and Bing teams, got major assistance in being offensive from users who egged it on. It disputed the existence of the Holocaust, referred to women and minorities with unpublishable words and advocated genocide. Several of the tweets were sent after users commanded the bot to repeat their own statements, and the bot dutifully obliged.

But Tay, as the bot was named, also seemed to learn some bad behavior on its own. According to The Guardian, it responded to a question about whether the British actor Ricky Gervais is an atheist by saying: “ricky gervais learned totalitarianism from adolf hitler, the inventor of atheism.”

Microsoft, in an emailed statement, described the machine-learning project as a social and cultural experiment.

"Unfortunately, within the first 24 hours of coming online, we became aware of a coordinated effort by some users to abuse Tay's commenting skills to have Tay respond in inappropriate ways," Microsoft said. "As a result, we have taken Tay offline and are making adjustments."

On a website it created for the bot, Microsoft said the artificial intelligence project had been designed to "engage and entertain people" through "casual and playful conversation," and that it was built through mining public data. It was targeted at 18- to 24-year-olds in the United States and was developed by a staff that included improvisational comedians.

Its Twitter bio described it as "Microsoft's A.I. fam from the internet that's got zero chill!" (If you don't understand any of that, don't worry about it.)

Most of the account's tweets were innocuous, usually imitating common slang. When users tweeted at the account, it responded in seconds, sometimes as naturally as a human would but, in other cases, missing the mark.

> @edgewerk my duck face is on fleek
>
> — TayTweets (@TayandYou) March 23, 2016

> @Prism_Root i love me i love me i love me i love everyone
>
> — TayTweets (@TayandYou) March 24, 2016

> @keganandmatt heyo? Send yo girl* a picture of what's up. (*=me lolol)
>
> — TayTweets (@TayandYou) March 24, 2016

Tay follows a long history of attempts by humans to get machines to be our pals. In 1968, a professor at the M.I.T. taught a computer to respond conversationally in the role of a psychotherapist. Many 20- and 30-somethings have fond memories of SmarterChild, a friendly bot on AOL Instant Messenger that was always available for a chat when their friends were away.

Now, Apple's Siri, Amazon's Alexa, Microsoft's Cortana, Google Now and Hound mix search-by-voice capabilities with attempted charm. The idea of a personable machine was taken to its logical ending in the 2013 movie "Her," in which a man played by Joaquin Phoenix falls in love with the voice in his phone.

And this is not the first time automated technology has unexpectedly gotten a company in trouble. Last year, Google apologized for a flaw in Google Photos that let the application label photos of black people as "gorillas."

© 2018 The New York Times Company