

LEGAL Drivers who switch to Allstate can save \$356.  Quote Now  Allstate

Subscribe

Latest Issues

SCIENTIFIC AMERICAN 

Cart 0

Sign In | Stay Informed

[CORONAVIRUS](#) [THE SCIENCES](#) [MIND](#) [HEALTH](#) [TECH](#) [SUSTAINABILITY](#) [VIDEO](#) [PODCASTS](#) [BLOGS](#) [PUBLICATIONS](#) [Q](#)



COMPUTING 60-SECOND SCIENCE

Artificial Intelligence Learns to Talk Back to Bigots

By Christopher Intagliata on October 10, 2019





Credit: Marcus Butt *Getty Images*

Algorithms are already used to remove online hate speech. Now scientists have taught an AI to respond—which they hope might spark more discourse. Christopher Intagliata reports.

[Full Transcript](#)

Social media platforms like Facebook use a combination of artificial intelligence and human moderators to scout out and eliminate hate speech. But now researchers have developed a new AI tool that wouldn't just scrub hate speech but would actually craft responses to it, like: "The language used is highly offensive. All ethnicities and social groups deserve tolerance."

"And this type of intervention response can hopefully short-circuit the hate cycles that we often get in these types of forums."

DIGITAL Workforce.ai  by Amelia 

Anna Bethke, a data scientist at Intel. The idea, she says, is to fight hate speech with *more* speech—an approach advocated by the ACLU and the U.N. High Commissioner for Human Rights.

So with her colleagues at U.C. Santa Barbara, Bethke got access to more than 5,000 conversations from the site Reddit and nearly 12,000 more from Gab—a social media site where many users banned by Twitter tend to resurface.

The researchers had real people craft sample responses to the hate speech in those Reddit and Gab conversations. Then they let natural-language-processing algorithms learn from the real human responses and craft their own, such as: “I don’t think using words that are sexist in nature contribute to a productive conversation.”

Which sounds pretty good. But the machines also spit out slightly head-scratching responses like this one: “This is not allowed and un time to treat people by their skin color.”

And when the scientists asked human reviewers to blindly choose between human responses and machine responses—well, most of the time, the humans won. The team published the results on the site *Arxiv* and will present them next month in Hong Kong at the [Conference on Empirical Methods in Natural Language Processing](#). [[Jing Qian et al., A benchmark dataset for learning to intervene in online hate speech](#)]

Ultimately, Bethke says, the idea is to spark more conversation.

“Not just to have this discussion between a person and a bot but to start to elicit the conversations within the communities themselves—between the people that might be being harmful and those they’re potentially harming.”

In other words, to bring back good ol’ civil discourse?

“Oh! I don’t know if I’d go that far. But it sort of sounds like that’s what I just proposed, huh?”

—Christopher Intagliata

[The above text is a transcript of this podcast.]

[Close Transcript](#)

[Rights & Permissions](#)

ABOUT THE AUTHOR(S)

Christopher Intagliata

Recent Articles

[Computers Confirm Beethoven’s Influence](#)

[Having an Albatross around Your Boat](#)

[Did Animal Calls Start in the Dark?](#)

Interview Amelia Today



Unlike chatbots, IPsoft’s cognitive agent Amelia does more than just recognize keywords.



ADVERTISEMENT

 **Flamingos Can Be Picky about Company**
60-Second Science - May 7, 2020 - By Jason G. Goldman | 03:00
[Full Transcript](#)




 **Horses Recognize Pics of Their Keepers**
60-Second Science - May 5, 2020 - By Susanne Bard | 03:00
[Full Transcript](#)



 **Tapirs Help Reforestation via Defecation**
60-Second Science - May 1, 2020 - By Jason G. Goldman | 03:01
[Full Transcript](#)



 **Virus-Infected Bees Practice Social Distancing**
60-Second Science - April 30, 2020 - By Karen Hopkin | 03:24
[Full Transcript](#)



 **New Data on Killer House Cats**
60-Second Science - April 29, 2020 - By Jason G. Goldman | 03:23
[Full Transcript](#)



 **Science News Briefs from around the World**
60-Second Science - April 28, 2020 - By Sarah Lewin Frasier | 01:48
[Full Transcript](#)



See all podcasts

Publishing research from all areas of
the natural sciences



Support Science Journalism

Discover world-changing science. Explore our digital archive
back to 1845, including articles by more than 150 Nobel Prize
winners.

Subscribe Now!



FOLLOW US



SCIENTIFIC AMERICAN ARABIC

العربية

[Return & Refund Policy](#)

[About](#)

[Press Room](#)

[FAQs](#)

[Contact Us](#)

[Site Map](#)

[Advertise](#)

[SA Custom Media](#)

[Terms of Use](#)

[Privacy Policy](#)

[Use of Cookies](#)

[International Editions](#)

Scientific American is part of Springer Nature, which owns or has commercial relations with thousands of scientific publications (many of them can be found at www.springernature.com/us).
Scientific American maintains a strict policy of editorial independence in reporting developments in science to our readers.

© 2020 SCIENTIFIC AMERICAN, A DIVISION OF SPRINGER NATURE AMERICA, INC.

ALL RIGHTS RESERVED.

