The New York Times | https://nyti.ms/2WF1apE

There Is a Racial Divide in Speech-Recognition Systems, Researchers Say

Technology from Amazon, Apple, Google, IBM and Microsoft misidentified 35 percent of words from people who were black. White people fared much better.

By Cade Metz

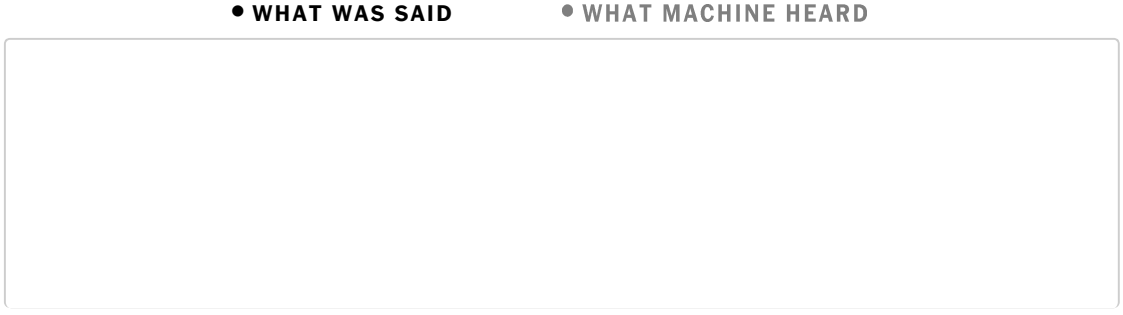
March 23, 2020

SAN FRANCISCO — With an iPhone, you can dictate a text message. Put Amazon's Alexa on your coffee table, and you can request a song from across the room.

But these devices may understand some voices better than others. Speech recognition systems from five of the world's biggest tech companies — Amazon, Apple, Google, IBM and Microsoft — make far fewer errors with users who are white than with users who are black, according to a study published Monday in the journal Proceedings of the National Academy of Sciences.

The systems misidentified words about 19 percent of the time with white people. With black people, mistakes jumped to 35 percent. About 2 percent of audio snippets from white people were considered unreadable by these systems, according to the study, which was conducted by researchers at Stanford University. That rose to 20 percent with black people.

Here's an audio clip of a 40-year-old black man speaking



> "I~~me~~ mean, I knew~~know~~ I'~~m~~ was kinda tall for high school~~asking~~. I didn't wanna play center. I didn't because center don't~~send it on~~ have the ball that much. You get the ball occasionally when you in the post, I mean, but I didn't want to play it."

The study, which took an unusually comprehensive approach to measuring bias in speech recognition systems, offers another cautionary sign for A.I. technologies rapidly moving into everyday life.

Other studies have shown that as facial recognition systems move into police departments and other government agencies, they can be far less accurate when trying to identify women and people of color. Separate tests have uncovered sexist and racist behavior in "chatbots," translation services, and other systems designed to process and mimic written and spoken language.

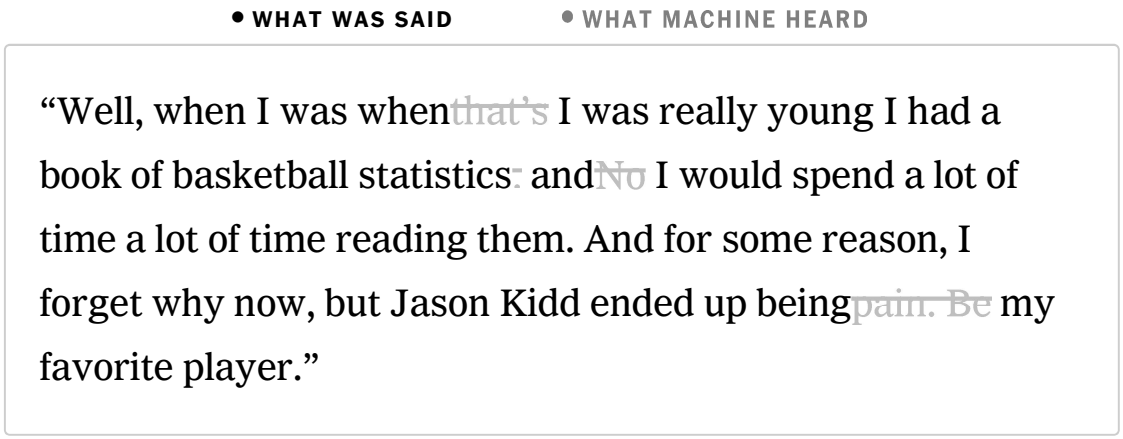
"I don't understand why there is not more due diligence from these companies before these technologies are released," said Ravi Shroff, a professor of statistics at New York University who explores bias and discrimination in new technologies. "I don't understand why we keep seeing these problems."

All these systems learn by analyzing vast amounts of data. Facial recognition systems, for instance, learn by identifying patterns in thousands of digital images of faces.

In many cases, the systems mimic the biases they find in the data, similar to children picking up bad habits from their parents. Chatbots, for example, learn by analyzing reams of human dialogue. If this dialogue associates women with housework and men with C.E.O. jobs, the chatbots will do the same.

The Stanford study indicated that leading speech recognition systems could be flawed because companies are training the technology on data that is not as diverse as it could be — learning their task mostly from white people, and relatively few black people.

Here's an audio clip of a 30-year-old white man



"Here are probably the five biggest companies doing speech recognition, and they are all making the same kind of mistake," said John Rickford, one of the Stanford researchers behind the study, who specializes in African-American speech. "The assumption is that all ethnic groups are well represented by these companies. But they are not."

The study tested five publicly available tools from Apple, Amazon, Google, IBM and Microsoft that anyone can use to build speech recognition services. These tools are not necessarily what Apple uses to build Siri or Amazon uses to build Alexa. But they may share underlying technology and practices with services like Siri and Alexa.

Each tool was tested last year, in late May and early June, and they may operate differently now. The study also points out that when the tools were tested, Apple's tool was set up differently from the others and required some additional engineering before it could be tested.

Apple and Microsoft declined to comment on the study. An Amazon spokeswomen pointed to a web page where the company says it is constantly improving its speech recognition services. IBM did not respond to requests for comment.

Justin Burr, a Google spokesman, said the company was committed to improving accuracy. "We've been working on the challenge of accurately recognizing variations of speech for several years, and will continue to do so," he said.

The researchers used these systems to transcribe interviews with 42 people who were white and 73 who were black. Then they compared the results from each group, showing a significantly higher error rate with the people who were black.

Craig Federighi, Apple's senior vice president of software engineering, spoke about Siri at a 2018 conference. Marcio Jose Sanchez/Associated Press

The best performing system, from Microsoft, misidentified about 15 percent of words from white people and 27 percent from black people. Apple's system, the lowest performer, failed 23 percent of the time with whites and 45 percent of the time with black people.

Based in a largely African-American rural community in eastern North Carolina, a midsize city in western New York and Washington, D.C., the black testers spoke in what linguists call African-American Vernacular English — a variety of English sometimes spoken by African-Americans in urban areas and other parts of the United States. The white people were in California, some in the state capital, Sacramento, and others from a rural and largely white area about 300 miles away.

The study found that the “race gap” was just as large when comparing the identical phrases uttered by both black and white people. This indicates that the problem lies in the way the systems are trained to recognize sound. The companies, it seems, are not training on enough data that represents African-American Vernacular English, according to the researchers.

Here’s an audio clip of a 40-year-old woman speaking in African-American Vernacular English

• WHAT WAS SAID • WHAT MACHINE HEARD

“My mom you know when she was had the store and she would used ~~will use~~ to close the ~~closest~~ store~~:~~ about~~By~~ eleven o-clock and she would promise us, now when we’re closed up~~and I were closer. We could~~ we’re just gonna~~go~~ take a ride and ~~around~~ look at the town.”

“The results are not isolated to one specific firm,” said Sharad Goel, a professor of engineering at Stanford and another researcher involved in the study. “We saw qualitatively similar patterns across all five firms.”

The companies are aware of the problem. In 2014, for example, Google researchers published a paper describing bias in an earlier breed of speech recognition.

In November, during a speech at Stanford dedicated to “ethical” artificial intelligence, Eric Schmidt, the former Google chief executive and chairman, said Google and the rest of Silicon Valley were well aware that the way A.I. systems were being built needed fixing.

“We know the data has bias in it. You don’t need to yell that as a new fact,” he said. “Humans have bias in them, our systems have bias in them. The question is: What do we do about it?”

A variety of consumer products are using speech-recognition technology like the Google Assistant. Joe Buglewicz for The New York Times

Companies like Google may have trouble gathering the right data, and they may not be motivated enough to gather it. “This is difficult to fix,” said Brendan O’Connor, a professor at the University of Massachusetts Amherst who specializes in A.I. technologies. “The data is hard to collect. You are fighting an uphill battle.”

The companies may face a chicken-and-egg problem. If their services are used mostly by white people, they will have trouble gathering data that can serve black people. And if they have trouble gathering this data, the services will continue to be used mostly by white people.

“Those feedback loops are kind of scary when you start thinking about them,” said Noah Smith, a professor at the University of Washington. “That is a major concern.”