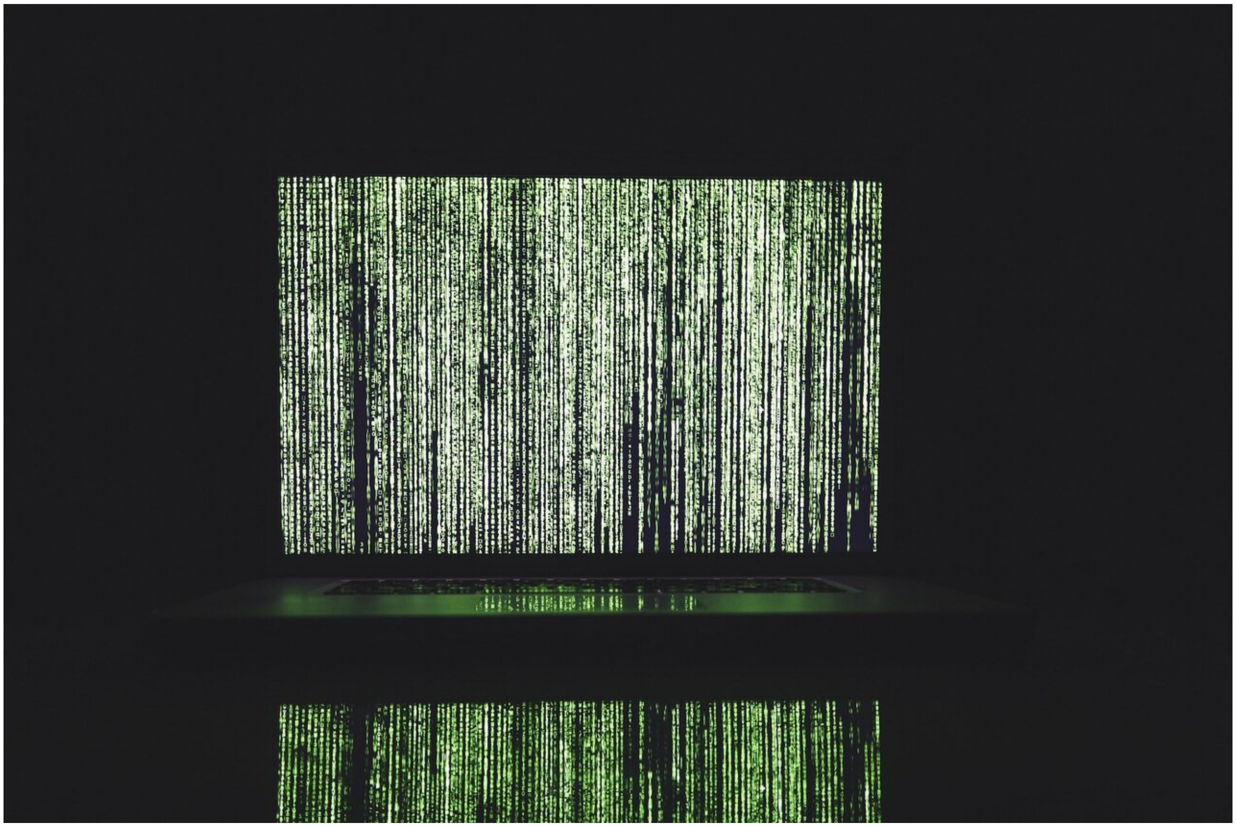
Even computer algorithms can be biased. Scientists have different ideas of how to prevent that

November 26 2019, by Amina Khan, Los Angeles Times



Credit: CC0 Public Domain

Scientists say they've developed a framework to make computer algorithms "safer" to use without creating bias based on race, gender or other factors. The trick, they say, is to make it possible for users to tell the algorithm what kinds of pitfalls to avoid—without having to know a lot about statistics or artificial intelligence.

With this safeguard in place, hospitals, companies and other potential users who may be wary of putting machine learning to use could find it a more palatable tool for helping them solve problems, according to a report in this week's edition of the journal *Science*.

Computer algorithms are used to make decisions in a range of settings, from courtrooms to schools to online shopping sites. The programs sort through huge amounts of data in search of useful patterns that can be applied to future decisions.

But researchers have been wrestling with a problem that's become increasingly difficult to ignore: Although the programs are automated, they often provide biased results.

For example, an algorithm used to determine prison sentences predicted higher recidivism rates for black defendants found guilty of crimes and a lower risk for white ones. Those predictions turned out to be wrong, according to a ProPublica analysis.

Biases like this often originate in the real world. An algorithm used to determine which patients were eligible for a health care coordination program was under-enrolling black patients largely because the code relied on real-world health spending data—and black patients had fewer dollars spent on them than whites did.

Even if the information itself is not biased, algorithms can still produce unfair or other "undesirable outcomes," said Philip Thomas, an artificial intelligence researcher at the University of Massachusetts Amherst and lead author of the new study.

Sorting out which processes might be driving those unfair outcomes, and then fixing them, can be an overwhelming task for doctors, hospitals or other potential users who just want a tool that will help them make better decisions.

"They're the experts in their field but perhaps not in machine learning—so we shouldn't expect them to have detailed knowledge of how algorithms work in order to control the behavior of the algorithms," Thomas said. "We want to give them a simple interface to define undesirable behavior for their application and then ensure that the algorithm will avoid that behavior with high probability."

So the computer scientists developed a different type of algorithm that allowed users to more easily define what bad behavior they wanted their program to avoid.

This, of course, makes the algorithm designers' job more difficult, Thomas said, because they have to build their algorithm without knowing what biases or other problematic behaviors the eventual user won't want in the program.

"Instead, they have to make the algorithm smart enough to understand what the user is saying is undesirable behavior, and then reason entirely on its own about what would cause this behavior, and then avoid it with high probability," he said. "That makes the algorithm a bit more complicated, but much easier for people to use responsibly."

To test their new framework, the researchers tried it out on a dataset of entrance exam scores for 43,303 Brazilian students and the grade point averages they earned during their first three semesters at college.

Standard algorithms that tried to predict a student's GPA based on his or her entrance exam scores were biased against women: The grades they predicted for women were lower than were actually the case, and the grades they predicted for men were higher. This caused an error gap between men and women that averaged 0.3 GPA points—enough to make a major difference in a student's admissions prospects.

The new algorithm, on the other hand, shrank that error range to within 0.05 GPA points—making it a much fairer predictor of students' success.

The computer scientists also tried out their framework on simulated data for diabetes patients. They found it could adjust a patient's insulin doses more effectively than a standard algorithm, resulting in far fewer unwanted episodes of hypoglycemia.

But others questioned the new approach.

Dr. Leo Anthony Celi, an intensivist at Beth Israel Deaconess Medical Center and research scientist at MIT, argued that the best way to avoid bias and other problems is to keep machine learning experts in the loop throughout the entire process rather than limiting their input to the initial design stages. That way they can see if an algorithm is behaving badly and make any necessary fixes.

"There's just no way around that," said Celi, who helped develop an artificial intelligence program to improve treatment strategies for patients with sepsis.

Likewise, front-line users such as doctors, nurses and pharmacists should take a more active role in the development of the algorithms they rely upon, he said.

The authors of the new study were quick to point out that their framework was more important than the algorithms they generated by using it.

"We're not saying these are the best algorithms," said Emma Brunskill, a computer scientist at Stanford University and the paper's senior author. "We're hoping that other researchers at their own labs will continue to make better algorithms."

Brunskill added that she'd like to see the new framework encourage people to apply algorithms to a broader range of health and social problems.

The new work is sure to stir up debate—and perhaps more needed conversations between the healthcare and machine learning communities, Celi said.

"If it makes people have more discussions then I think it's valuable," he said.

©2019 Los Angeles Times
Distributed by Tribune Content Agency, LLC.

Citation: Even computer algorithms can be biased. Scientists have different ideas of how to prevent that (2019, November 26) retrieved 23 April 2024 from https://techxplore.com/news/2019-11-algorithms-biased-scientists-ideas.html