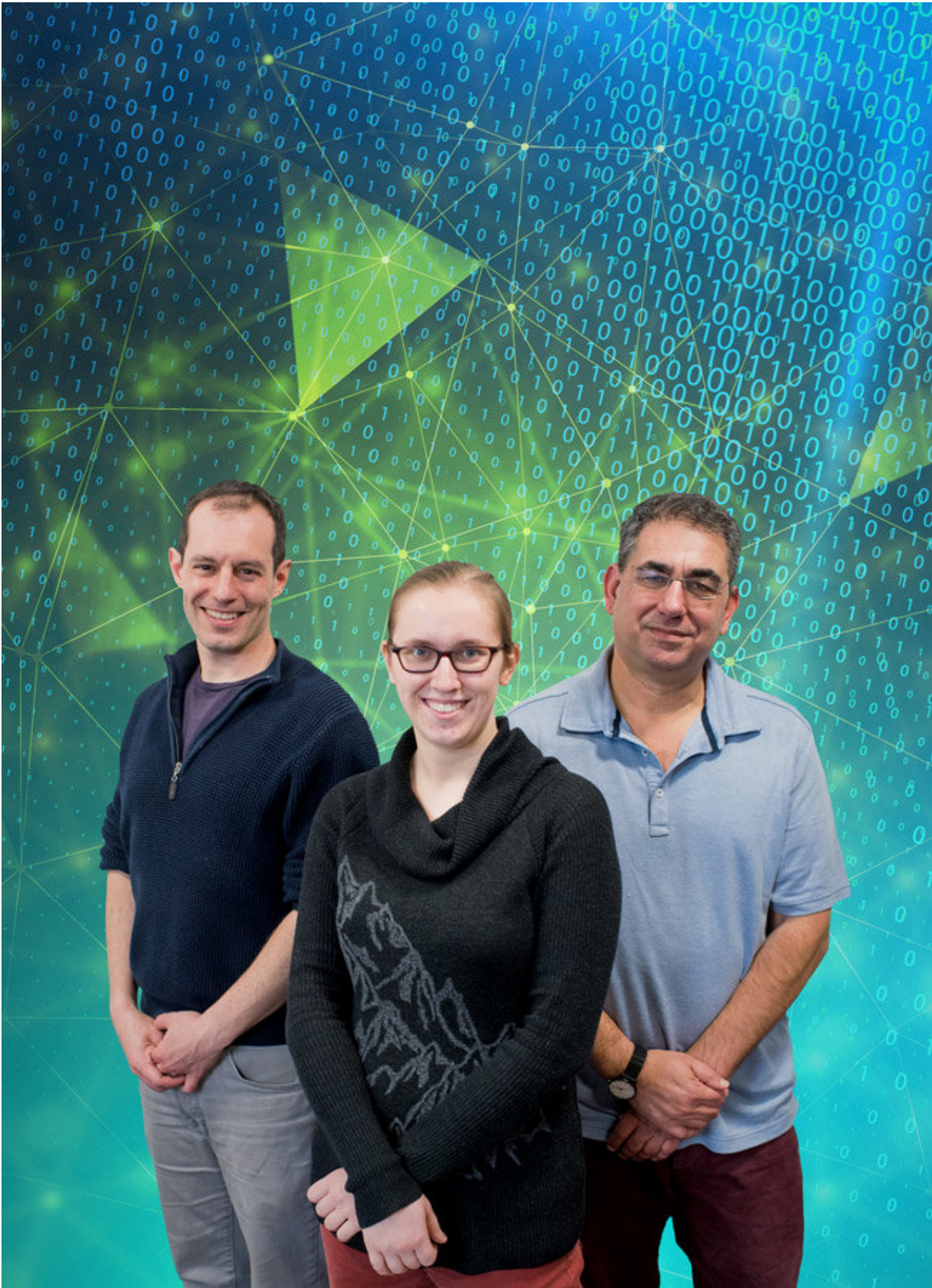


Are computer-aided decisions actually fair?

December 14 2018, by Art Jahnke



(From left) Adam Smith, Sarah Scheffler, and Ran Canetti. Credit: Jackie Ricciardi

Algorithmic fairness is increasingly important because as more decisions of greater importance are made by computer programs, the potential for harm grows. Today, algorithms are already widely used to determine credit scores, which can mean the difference between owning a home and renting one. And they are used in predictive policing, which suggests a likelihood that a crime will be committed, and in scoring how likely a criminal will commit another crime in the future, which influences the severity of sentencing.

That's a problem, says Adam Smith, a professor of computer science at Boston University, because the design of many algorithms is far from transparent.

"A lot of these systems are designed by [private companies](#) and their details are proprietary," says Smith, who is also a data science faculty fellow at the Hariri Institute for Computing. "It's hard to know what they are doing and who is responsible for the decisions they make."

Recently, Smith and a joint team of BU-MIT computer scientists reexamined this problem, hoping to learn what, if anything, can be done to understand and minimize bias from decision-making systems that depend on [computer programs](#).

The BU researchers—Smith, Ran Canetti, a professor of computer science and director of the Hariri Institute's Center for Reliable Information Systems and Cyber Security, and Sarah Scheffler (GRS'21), a computer science doctoral candidate—are working with MIT Ph.D.

students Aloni Cohen, Nishanth Dikkala, and Govind Ramnarayan to design systems whose decisions about all subsets of the population are equally accurate.

Their work was recently accepted for publication at the upcoming 2019 Association for Computing Machinery conference on Fairness, Accountability, and Transparency, nicknamed "ACM FAT."

The researchers believe that a [system](#) that discriminates against people who have had a hard time establishing a credit history will perpetuate that difficulty, limiting opportunity for a subset of the population and preserving existing inequalities. What that means, they say, is that automated ranking systems can easily become self-fulfilling prophecies, whether they are ranking the likelihood of default on a mortgage or the quality of a university education.

"Automated systems are increasingly complex, and they are often hard to understand for lay people and for the people about whom decisions are being made," Smith says.

The problem of self-fulfilling predictions

"The interaction between the [algorithm](#) and [human behavior](#) is such that if you create an algorithm and let it run, it can create a different society because humans interact with it," says Canetti. "So you have to be very careful how you design the algorithm."

That problem, the researchers say, will get worse as future algorithms use more outputs from past algorithms as inputs.

"Once you've got the same computer program making lots of decisions, any biases that exist are reproduced many times over on a larger scale," Smith says. "You get the potential for a broad societal shift caused by a

computer program."

But how exactly can an algorithm, which is basically a mathematical function, be biased?

Scheffler suggests two ways: "One way is with biased data," she says. "If your algorithm is based on historical data, it will soon learn that a particular institution prefers to accept men over women. Another way is that there are different accuracies on different parts of the population, so maybe an algorithm is really good at figuring out if [white people](#) deserve a loan, but it could have high error rate for people who are not white. It could have 90 percent accuracy on one set of the population and 50 percent on another set."

"That's what we are looking at," says Smith. "We're asking 'How is the system making mistakes?' and 'How are these mistakes spread across different parts of the population?'"

The real-world impact of algorithmic bias

In May 2016, reporters from ProPublica, a nonprofit investigative newsroom, examined the accuracy of COMPAS, one of several algorithmic tools used by court systems to predict recidivism, or the likelihood that a criminal defendant will commit another crime. The initial findings were not reassuring.

When ProPublica researchers compared the tool's predicted risk of recidivism with actual recidivism rates over the following two years, they found that, in general, COMPAS got things right 61 percent of the time. They also found that predictions of violent recidivism were correct only 20 percent of the time.

More troubling, they found that black defendants were far more likely

than white defendants to be incorrectly deemed more likely to commit crime again, and white defendants were more likely than black defendants to be incorrectly deemed low risk to recidivate. According to ProPublica's article, this was a clear demonstration of bias by the algorithm.

In response, Northpointe Inc., the creator of COMPAS, published another study that argued that the COMPAS algorithm is in fact fair according to a different statistical measure of bias: calibration.

Northpointe's software is widely used, and like many algorithmic tools, its calculations are proprietary, but the company did tell ProPublica that its formula for predicting who will recidivate is derived from answers to 137 questions whose answers come either from defendants or from criminal records.

Northpointe's study found that for each risk score, the fraction of white defendants who received this score and recidivated (out of all white defendants who received this score) roughly equals the fraction of black defendants who received this score and recidivated, out of all black defendants who received this score.

"ProPublica and Northpointe came to different conclusions in their analyses of COMPAS' fairness. However, both of their methods were mathematically sound—the opposition lay in their different definitions of fairness," Scheffler says.

The bottom line is that any imperfect prediction mechanism (either algorithmic or human) will be biased according to at least one of the two approaches: the error-balancing approach used by ProPublica, and the calibration method favored by Northpointe.

Overcoming algorithmic bias

When it came to solving the problem of algorithmic bias, the BU-MIT research team created a method of identifying the subset of the population that the system fails to judge fairly, and sending their review to a different system that is less likely to be biased. That separation guarantees that the method errs in more balanced ways regarding the individuals for whom it does make a decision.

And while the researchers found many situations where that solution appeared to work well, they remain concerned about how the different systems would work together. "There are many different measures of fairness," says Scheffler, "and there are trade-offs between them. So to what extent are the two systems compatible with the notion of fairness we want to achieve?"

"What happens to those people whose decisions would be deferred really influences how we view the system as a whole," says Smith. "At this point, we are still wrapping our heads around what the different solutions would mean."

Still, says Canetti, the research points to a possible way out of the statistical bias conundrum, one that could enable the design of algorithms that minimize the bias. That challenge, he says, will require expertise from many disciplines.

Provided by Boston University

Citation: Are computer-aided decisions actually fair? (2018, December 14) retrieved 18 May 2024 from <https://phys.org/news/2018-12-computer-aided-decisions-fair.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.