

NEWS &amp; TECHNOLOGY 29 March 2017

# Bias test to prevent algorithms discriminating unfairly



Algorithms discriminate, too  
Frances Roberts/Alamy Stock Photo

By **Matt Reynolds**

COMPUTERS are getting ethical. A new approach for testing whether algorithms contain hidden biases aims to prevent automated systems from perpetuating human discrimination.

Machine learning is increasingly being used to make sensitive decisions, says Matt Kusner at the Alan Turing Institute in London. In some US states, judges make sentencing decisions and set bail conditions using algorithms that calculate the likelihood that someone will reoffend. Other algorithms assess whether a person should be offered a loan or a job interview.

But it is often unclear how these systems come to their conclusions, which makes it impossible to tell if they are fair ones. An algorithm might conclude that people from a certain demographic are less likely to pay back a loan, for example, if it is trained on a data set in which loans were unfairly distributed in the first place. “In machine learning, we have this problem of racism in and racism out,” says Chris Russell, also at the Alan Turing Institute.

Kusner and Russell are part of a team that has developed a framework to identify and eliminate algorithmic bias. A fair algorithm, the team says, is one that makes the same decision about an individual regardless of demographic background. Because someone’s background can influence other aspects of their lives, such as their neighbourhood or academic achievement, it’s not enough to ignore obvious race or gender cues; you have to consider how variables interact.

So the team maps out variables in a data set and tests how they might skew decision-making processes. If there is evidence of bias, the researchers find a way to remove or compensate for it.

The team applied this to 2014 stop-and-frisk data from the New York City Police Department, modelling variables that influenced police officers’ decisions to stop someone.

First, the team considered all variables, including the skin colour and appearance of stopped people. Then, they considered only data points related to actual criminality, such as whether the person was found to be carrying a weapon and whether they were arrested.

The team found that police generally saw black and Hispanic men as more criminal than they did white men, stopping them more often – so a machine-learning analysis of these events might deduce that criminality is correlated with skin colour.

But evidence of actual criminality was similar across the different groups. A model trained like this could help decide where to place police patrols without discriminating by ethnic origin ([arxiv.org/abs/1703.06856](https://arxiv.org/abs/1703.06856)).

“I think these guys get it right,” says Simon DeDeo at Carnegie Mellon University in Pittsburgh, Pennsylvania. Rather than trusting algorithms to make ethical choices, he says, we need to work out what is fair. “The big problem is getting the computer to be able to explain itself to us.”

The team now wants to try the approach in other areas. It could be useful for firms that are legally obliged to keep their processes free from discrimination, such as mortgage providers.

*This article appeared in print under the headline “Bias test to keep algorithms ethical”*

---

Magazine issue 3119, published 1 April 2017



**Head of EHS**  
[Apply for this job](#)



**Production Planner**  
[Apply for this job](#)



**QA Specialist - Systems**  
[Apply for this job](#)



**Antibody Discovery Scientist -  
Hybridoma (Sci I)**  
[Apply for this job](#)

**More jobs ►**