

Exploring the use of algorithms in the criminal justice system

31 May 2017, by Vignesh Ramachandran



Using an algorithm to help set bail could keep many defendants out of jail without increasing the risk of flight to avoid trial. Credit: Stocksy/Clisque Images

Courts nationwide are making greater use of computer algorithms to help determine whether defendants should be released into the community while they await trial. The idea is to accurately determine whether a defendant poses a flight risk or a danger to the community, and reduce the potential for human bias.

San Francisco and Chicago are among the jurisdictions actively using pretrial risk assessment algorithms, says Sharad Goel, an assistant professor of management science and engineering.

Goel has studied the use of algorithms [in complex decision making](#) and examined [the fairness of such tools](#). As co-author of a recent Harvard Business Review [article](#), Goel argues that simple, statistically informed decision rules can dramatically improve judicial determinations. But, he cautions, algorithms are not a complete fix.

"Algorithms are good at narrowly estimating risk, but they can't set policy," Goel says. "They can't tell you how many people to detain, or whether we should end money bail altogether, as some cities

have done. They can't tell you how much to invest in pretrial services or what those services should be. And they can't incorporate every factor in every case, so we still need humans to make the final decision."

We recently met with Goel to discuss some of the benefits and complexities of using these algorithms. Excerpts:

How are algorithms used in the criminal justice system to make pretrial release decisions?

Algorithms are mostly used in two ways: to estimate a [defendant's](#) flight risk, and to assess his or her threat to public safety. For example, based on a variety of factors, like age and criminal history, these algorithms rate a defendant's likelihood to re-offend, usually on a scale from 1 to 10. Judges use these risk scores to help decide which defendants to release and which to detain pending trial.

Why use algorithms? Why not rely on human judges alone?

Computers are good at estimating the likelihood of an event given structured information, like a defendant's criminal history. Algorithms can pick out which pieces of information matter and which should be ignored to generate accurate estimates of risk.

In theory, judges try to do the same thing, but it's easy for people to focus on the wrong factors and let implicit biases creep in. And some judges are just tougher than others, so there isn't a consistent standard. If you're assigned to a strict judge rather than a lenient one, you might get different results.

We've looked at over 100,000 judicial decisions. By using an [algorithm](#), we find that you could detain half as many defendants without increasing the number who fail to appear at trial. A lot of people who pose very little risk are being needlessly

detained. There's a huge social and financial cost to that. significant shortcomings, which we and others have pointed out.

What does it mean for an algorithm to be fair?

Defining [fairness](#) is a complicated and still open problem. I doubt we'll ever reach consensus, but there are a few common ways to think about it.

Some say an algorithm is fair if it doesn't consider sensitive attributes, like race or gender. But even if you don't explicitly consider such attributes, that information is usually baked into other factors, like place of residence or income. Yale law professor Ian Ayres has persuasively [argued](#) that in some situations it's even unfair not to consider race when making decisions.

Others say an algorithm is fair only if its impact is the same on all race groups. For example, if more blacks than whites are rated high-risk by the algorithm, people in this camp would call that unfair.

Related to this idea of impact, some define fairness in terms of error rates. Algorithms seek to predict which defendants are most likely to commit new offenses, or to "recidivate." One can look back at these predictions and ask how often the algorithms were wrong: How often did the algorithm classify blacks and whites as high-risk of re-offending when in fact those defendants did not go on to commit any new crimes? If black non-recidivists are more likely to be classified as high-risk than white non-recidivists, that would be unfair by this measure.

How do you define fair?

My preferred definition of fairness is that equally risky defendants are treated equally, regardless of race. For example, if the available information indicates that a white defendant and a black defendant both have a 30 percent chance of committing a violent crime, both defendants are either released or both are detained. To me this definition makes intuitive sense, and we [show](#) that there are strong legal and policy arguments supporting it.

The other popular definitions of fairness have

For instance, consider an algorithm that disproportionately classifies black defendants as high-risk. I wouldn't automatically call such disparate impacts unfair. For a variety of complex social and economic reasons, black defendants on average might be riskier than whites, in which case we would expect detention rates to reflect those differences.

The same is true for disparate error rates in estimating recidivism. In our [paper](#), we look at data from Florida and find that it's objectively harder to correctly classify black defendants than white defendants. That's because a disproportionate number of black defendants have about even odds of reoffending, based on their prior criminal records. These defendants are not clearly going to commit a crime, but also are not clearly not going to commit a crime. Because it's hard to predict the behavior of such defendants, that drives up error rates for blacks as a group. As with unequal detention rates, I wouldn't call unequal error rates inherently unfair.

In time, if socioeconomic disparities narrow between whites and blacks, I suspect that an algorithm which is fair by one measure will be fair by the others. But while race differences persist, these definitions are largely incompatible. That tension is at the heart of recent debates over what makes an algorithm fair.

What are some misconceptions about the fairness of algorithms?

Perhaps the biggest misconception is that we should worry more about decisions made by algorithms than those made by humans. Many of the fairness issues ascribed to algorithms apply equally to human judges. And some problems, like inconsistency, afflict humans more than computers.

It's also common to conflate disparate impact with discrimination. We've [argued](#) that algorithms which many people, including legal experts, would consider fair necessarily lead to racial disparities. It's easy to latch onto these disparities as evidence of bias, but that misses the complexity of the

problem.

Another big misconception is that algorithms are inherently unfair because they are based on imperfect data. Bad data is a serious issue which we shouldn't ignore, but algorithms and humans can only use the information that's available. Fairness must be viewed in context.

So should the criminal justice system be using algorithms, and if so, how and to what extent? Where is this headed in the future?

Algorithms will almost certainly play an increasingly prominent role in criminal justice. In cities where pretrial risk assessment tools have been deployed, fewer defendants are detained with little to no decrease in public safety.

To gain wider support and adoption, I think these algorithms need to be developed with more transparency. The leading risk assessment tools are often built under a veil of secrecy, which understandably sows misunderstanding and distrust.

Algorithms have important limitations, but they can also dramatically improve the equity of decisions in our [criminal justice](#) system.

Provided by Stanford University

APA citation: Exploring the use of algorithms in the criminal justice system (2017, May 31) retrieved 21 April 2021 from <https://phys.org/news/2017-05-exploring-algorithms-criminal-justice.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.