Opinion | OP-ED CONTRIBUTOR

When an Algorithm Helps Send You to Prison

By ELLORA THADANEY ISRANI   OCT. 26, 2017

In 2013, police officers in Wisconsin arrested a man driving a car that had been used in a recent shooting. The man, Eric Loomis, pleaded guilty to attempting to flee an officer, and no contest to operating a vehicle without the owner's consent. Neither of his crimes mandates prison time.

At Mr. Loomis's sentencing, the judge cited, among other factors, Mr. Loomis's high risk of recidivism as predicted by a computer program called COMPAS, a risk assessment algorithm used by the state of Wisconsin. The judge denied probation and prescribed an 11-year sentence: six years in prison, plus five years of extended supervision.

No one knows exactly how COMPAS works; its manufacturer refuses to disclose the proprietary algorithm. We only know the final risk assessment score it spits out, which judges may consider at sentencing.

Mr. Loomis challenged the use of an algorithm as a violation of his due process rights to be sentenced individually, and without consideration of impermissible factors like gender. The Wisconsin Supreme Court rejected his challenge. In June, the United States Supreme Court declined to hear his case, meaning a majority of justices effectively condoned the algorithm's use. Their decision will have far-ranging effects.

Something about this story is fundamentally wrong: Why are we allowing a computer program, into which no one in the criminal justice system has any

insight, to play a role in sending a man to prison?

At a sentencing, it is entirely a judge's prerogative to prescribe a sentence within statutory guidelines. The obvious flaw with this system is bias. Judges might abuse this unchecked power to sentence based not only on relevant factors, such as the seriousness of a defendant's offense, but also on those that are morally and constitutionally problematic — for example, gender and race.

This is precisely why states are abdicating the responsibility for sentencing to a computer. Use of a computerized risk assessment tool somewhere in the criminal justice process is widespread across the United States, and some states, such as Colorado, even require it. States trust that even if they cannot themselves unpack proprietary algorithms, computers will be less biased than even the most well-meaning humans.

But shifting the sentencing responsibility to a computer does not necessarily eliminate bias; it delegates and often compounds it.

Algorithms like COMPAS simply mimic the data with which we train them. COMPAS's authors presumably fed historical recidivism data into their system. From that, the program ascertained what factors tend to make a defendant a higher risk. It then applied the patterns it gleaned to subsequent defendants, like Mr. Loomis, to spit out sentences that comport with existing trends.

But an algorithm that accurately reflects our world also necessarily reflects our biases. A ProPublica study found that COMPAS predicts black defendants will have higher risks of recidivism than they actually do, while white defendants are predicted to have lower rates than they actually do. (Northpointe Inc., the company that produces the algorithm, disputes this analysis.) The computer is *worse* than the human. It is not simply parroting back to us our own biases, it is exacerbating them.

Even if you think Mr. Loomis's sentencing procedure arrived at the appropriate result, the potential that the process the state took to arrive there was biased — in ways neither judges nor defendants nor prosecutors know — should alarm anyone.

Machine learning algorithms often work on a feedback loop. If they are not constantly retrained, they "lean in" to the assumed correctness of their initial determinations, drifting away from both reality and fairness. As a former Silicon Valley software engineer, I saw this time and again: Google's image classification algorithms mistakenly labeling black people as gorillas, or Microsoft's Twitter bot immediately becoming a "racist jerk."

Algorithms also lack the human ability to individualize. A computer cannot look a defendant in the eye, account for a troubled childhood or disability, and recommend a rehabilitative sentence. This is precisely the argument against mandatory minimum sentences — they rob judges of the discretion to deliver individualized justice — and it is equally cogent against machine sentencing.

For example, algorithms are often programmed to assume unidirectional causation: If A, then B. But is it truly that defendants with higher rates of recidivism warrant longer sentences or is it that defendants with longer sentences are kept out of their communities, unemployed and away from their families longer, naturally increasing their recidivism risk? A judge could investigate this nuance.

With transparency and accountability, algorithms in the criminal justice system do have potential for good. For example, New Jersey used a risk assessment program known as the Public Safety Assessment to reform its bail system this year, leading to a 16 percent decrease in its pre-trial jail population. The same algorithm helped Lucas County, Ohio double the number of pre-trial releases without bail, and cut pre-trial crime in half. But that program's functioning was detailed in a published report, allowing those with subject-matter expertise to confirm that morally troubling (and constitutionally impermissible) variables — such as race, gender and variables that could proxy the two (for example, ZIP code) — were not being considered.

For now, the only people with visibility into COMPAS's functioning are its programmers, who are in many ways less equipped than judges to deliver justice. Judges have legal training, are bound by ethical oaths, and must account for not only their decisions but also their reasoning in published opinions. Programmers lack each of these safeguards.

Computers may be intelligent, but they are not wise. Everything they know, we taught them, and we taught them our biases. They are not going to un-learn them without transparency and corrective action by humans.

Ellora Thadaney Israni (@elloraisrani) is a J.D. candidate at Harvard Law School and former software engineer at Facebook.

Follow The New York Times Opinion section on Facebook and Twitter (@NYTopinion), and sign up for the Opinion Today newsletter.