An Algorithm That Grants Freedom, or Takes It Away

Across the United States and Europe, software is making probation decisions and predicting whether teens will commit crime. Opponents want more human oversight.

By **Cade Metz** and **Adam Satariano**

Published Feb. 6, 2020   Updated Feb. 7, 2020

PHILADELPHIA — Darnell Gates sat at a long table in a downtown Philadelphia office building. He wore a black T-shirt with "California" in bright yellow letters on the chest. He had never been to the state, but he hoped to visit family there after finishing his probation.

When Mr. Gates was released from jail in 2018 — he had served time for running a car into a house in 2013 and later for violently threatening his former domestic partner — he was required to visit a probation office once a week after he had been deemed "high risk."

He called the visits his "tail" and his "leash." Eventually, his leash was stretched to every two weeks. Later, it became a month. Mr. Gates wasn't told why. He complained that conversations with his probation officers were cold and impersonal. They rarely took the time to understand his rehabilitation.

He didn't realize that an algorithm had tagged him high risk until he was told about it during an interview with The New York Times.

"What do you mean?" Mr. Gates, 30, asked. "You mean to tell me I'm dealing with all this because of a computer?"

In Philadelphia, an algorithm created by a professor at the University of Pennsylvania has helped dictate the experience of probationers for at least five years.

The algorithm is one of many making decisions about people's lives in the United States and Europe. Local authorities use so-called predictive algorithms to set police patrols, prison sentences and probation rules. In the Netherlands, an algorithm flagged welfare fraud risks. A British city rates which teenagers are most likely to become criminals.

Nearly every state in America has turned to this new sort of governance algorithm, according to the Electronic Privacy Information Center, a nonprofit dedicated to digital rights. Algorithm Watch, a watchdog in Berlin, has identified similar programs in at least 16 European countries.

As the practice spreads into new places and new parts of government, United Nations investigators, civil rights lawyers, labor unions and community organizers have been pushing back.

They are angered by a growing dependence on automated systems that are taking humans and transparency out of the process. It is often not clear how the systems are making their decisions. Is gender a factor? Age? ZIP code? It's hard to say, since many states and countries have few rules requiring that algorithm-makers disclose their formulas.

They also worry that the biases — involving race, class and geography — of the people who create the algorithms are being baked into these systems, as ProPublica has reported. In San Jose, Calif., where an algorithm is used during arraignment hearings, an organization called Silicon Valley De-Bug interviews the family of each defendant,

takes this personal information to each hearing and shares it with defenders as a kind of counterbalance to algorithms.

Two community organizers, the Media Mobilizing Project in Philadelphia and MediaJustice in Oakland, Calif., recently compiled a nationwide database of prediction algorithms. And Community Justice Exchange, a national organization that supports community organizers, is distributing a 50-page guide that advises organizers on how to confront the use of algorithms.

The algorithms are supposed to reduce the burden on understaffed agencies, cut government costs and — ideally — remove human bias. Opponents say governments haven't shown much interest in learning what it means to take humans out of the decision making. A recent United Nations report warned that governments risked "stumbling zombie-like into a digital-welfare dystopia."

Last year, Idaho passed a law specifying that the methods and data used in bail algorithms must be publicly available so the general public can understand how they work. In the Netherlands, a district court ruled Wednesday that the country's welfare-fraud software violated European human rights law, one of the first rulings against a government's use of predictive algorithms.

"Where is my human interaction?" Mr. Gates asked, sitting next to his lawyer in the boardroom of the Philadelphia public defender's office. "How do you win against a computer that is built to stop you? How do you stop something that predetermines your fate?"

'We walked into a hornet's nest'

State Representative Todd Stephens is a member of a Pennsylvania commission working to adopt predictive technology. Bastiaan Slabbers/Sipa USA, via Associated Press

On a recent Thursday, Todd Stephens sat in a food court across a street from Citizens Bank Park, home of the Philadelphia Phillies. He was explaining the latest effort to remake state sentencing practices with a predictive algorithm.

Predictive algorithms, at their most basic, work by using historical data to calculate a probability of future events, similar to how a sports book determines odds for a game or pollsters forecast an election result.

The technology builds on statistical techniques that have been used for decades, often for determining risk. They have been supercharged thanks to increases in affordable computing power and available data.

The private sector uses such tools all the time, to predict how likely people are to default on a loan, get sick or be in a car wreck, or whether they will click on an internet ad. Governments, which hold vast amounts of data about the public, have been eager to tap their potential.

A Republican member of the Pennsylvania House of Representatives, Mr. Stephens is part of a state commission working to adopt the technology. Like many states, Pennsylvania has mandated that an algorithm be developed to help courts decide the sentence after someone is convicted.

The idea, Mr. Stephens said, was to predict how likely people were to commit another crime by collecting information about them and comparing that to statistics describing known offenders. That might include age, sex, and past and current convictions.

The commission had proposed a plan that would have leaned on information provided by county probation departments. But the American Civil Liberties Union and community groups protested this plan during public meetings last fall. They worried it would expand the power of predictive algorithms used for probation, including the one that tagged Mr. Gates.

"We walked into a hornet's nest I didn't even know existed," Mr. Stephens said.

In response to the protests, the state commission recommended a much simpler setup based on software already used by the state courts. But even this algorithm is difficult for a layperson to understand. Asked to explain it, Mr. Stephens suggested speaking with another commissioner.

Nyssa Taylor, criminal justice policy counsel with the Philadelphia A.C.L.U., was among the protesters. She worries that algorithms will exacerbate rather than reduce racial bias. Even if governments share how the systems arrive at their decisions — which happens in Philadelphia in some cases — the math is sometimes too complex for most people to wrap their heads around.

Various algorithms embraced by the Philadelphia criminal justice system were designed by Richard Berk, a professor of criminology and statistics at Penn. These algorithms do not use ZIP codes or other location data that could be a proxy for race, he said. And though he acknowledged that a layperson couldn't easily understand the algorithm's decisions, he said human judgment has the same problem.

"All machine-learning algorithms are black boxes, but the human brain is also a black box," Dr. Berk said. "If a judge decides they are going to put you away for 20 years, that is a black box."

Mark Houldin, a former public defender who was also among the protesters, said he was concerned that the algorithms were unfairly attaching labels to individuals as they moved through the criminal justice system.

In an affidavit included with a lawsuit recently filed by the defender's office, a former Philadelphia probation officer said the probation department's predictive algorithm also affected arraignment hearings. For years, she said, if someone was arrested and charged with a new crime while on probation — and had been deemed "high risk" by the algorithm — the probation office automatically instructed the jail not to release the person.

A spokesman for Philadelphia County denied that the system operated this way. "Every detainer issued is reviewed by supervisory staff of" the probation department, he said, "and notice is sent to the appropriate judicial authority."

Mr. Houldin argued that the use of the probation algorithm in this situation would deny due process. "If you are arrested for a new crime, the presumption of innocence is gone," he said.

Dr. Berk, who designed the probation algorithm, said it was not designed to be used this way.

"One of things that I make really clear about this algorithm — and all others — is that they are hand tailored to a particular decision," he said. "If you move them to another decision, the warranty doesn't apply anymore."

Looking for welfare fraud in Rotterdam

Last year in Rotterdam, a rumor circulating in two predominantly low-income and immigrant neighborhoods claimed that the city government had begun using an experimental algorithm to catch citizens who were committing welfare and tax fraud.

Mohamed Saidi learned about it from a WhatsApp message that he initially thought was a joke. Mohamed Bouchkhachakhe first heard from his mother, who had been told by a friend. Driss Tabghi got word from a local union official.

The rumor turned out to be true.

The Dutch program, System Risk Indication, scans data from different government authorities to flag people who may be claiming unemployment aid when they are working, or a housing subsidy for living alone when they are living with several others.

The agency that runs the program, the Ministry of Social Affairs and Employment, said the data could include income, debt, education, property, rent, car ownership, home address and the welfare benefits received for children, housing and health care.

The algorithm produces "risk reports" on individuals who should be questioned by investigators. In Rotterdam, where the system was most recently used, 1,263 risk reports were produced in two neighborhoods.

"You're putting me in a system that I didn't even know existed," said Mr. Bouchkhachakhe, who works for a logistics company.

The program has been cloaked in secrecy. Even those who land on the list aren't informed. They aren't told how the algorithm is making its decisions, or given ways to appeal. In 2019, a city council hearing with the social ministry abruptly ended when members of the city government wouldn't sign nondisclosure agreements before receiving a briefing about how the system works.

Such disclosure would "interfere with the ability to effectively investigate," the ministry said in response to questions.

In a report in October, the United Nations special rapporteur on extreme poverty and human rights criticized the Dutch system for creating a "digital welfare state" that turns crucial decisions about people's lives over to machines.

"Whole neighborhoods are deemed suspect and are made subject to special scrutiny, which is the digital equivalent of fraud inspectors knocking on every door in a certain area," the report said. "No such scrutiny is applied to those living in better off areas."

Similar programs exist elsewhere. In North Carolina, IBM software has been used to identify Medicaid fraud. In London, local councils tested software to identify those who may be wrongly claiming a housing benefit. Systems are used to flag children who may be at risk of abuse.

In Rotterdam, opposition built after word about the techniques spread. Privacy rights groups, civil rights lawyers and the largest national labor union rallied citizens to fight the effort.

"They will not tell you if you are on the register," said Tijmen Wisman, an assistant professor of privacy law who runs a Dutch privacy group. He helped organize a meeting for roughly 75 residents in the affected neighborhoods, many taking video on their smartphones to share with neighbors.

The district court that sided with the opponents ordered an immediate halt to the risk algorithm's use. In the closely watched case, which is seen as setting a precedent in Europe about government use of predictive algorithms, the court said that the welfare program lacked privacy safeguards and that the government was inadequately transparent about how it worked. The decision can be appealed.

"The right to receive social security is being made conditional on exposing yourself to state surveillance," said Christiaan van Veen, a lawyer who is a special adviser on new technology and human rights for the United Nations.

Spotting teen trouble in Bristol, England

In areas dealing with years of budget cuts, algorithms present a way to help make up for reduced social services. The technology, officials say, helps them do more with less and identify people who may otherwise slip through the cracks.

Once a week in Bristol, England, a team gathers in a conference room to review the latest results from an algorithm meant to identify the most at-risk youths in the city and review caseloads. Representatives from the police and children's services and a member of the team that designed the software typically attend to scan the list of names.

With youth violence and crime on the rise, and with many youth programs and community centers where young people gathered having been closed, the local government turned to software to help identify children most in need. Officials there say the work provides evidence the technology can work if coupled with a human touch.

Last year, Bristol introduced a program that creates a risk score based on data pulled from police reports, social benefits and other government records. The system tallies crime data, housing information and any known links to others with high risk scores, and if the youth's parents were involved in a domestic incident. Schools feed in attendance records.

"You can get quite a complete picture," said Tom Fowler, 29, the data scientist who helped create the youth scoring system for the Bristol government.

The scores fluctuate depending on how recently the youths had an incident like a school suspension. The goal at the weekly meetings is to identify children at risk of being recruited into crime.

There's evidence that the algorithm identifies the right people, but the city is still figuring out how to translate the data into action. Last year, a teenager who had one of the highest risk scores stabbed someone to death. In a review of the killing, city officials concluded that they had taken the right steps. Mr. Fowler said a person can't be arrested simply because of the algorithm.

"He had a social worker and one-on-one coaching," said Mr. Fowler, who now works for a data-analytics company. "He made a really bad decision.

"You can't control for that. Data can only go so far. But it's pretty much the worst thing that can happen. That makes you do a bit of soul searching whether you did everything you could."

Dozens of local governments across Britain are turning to algorithms to guide their decision making, according to a 2018 investigation by the privacy group Big Brother Watch. The Guardian reported that one in three local councils used algorithms in some capacity for government programs.

In Bristol, the government has been open with the public about the program, posting some details online and holding community events. Opponents say it still isn't fully transparent. The young people and their parents do not know if they are on the list or given a way to contest their inclusion.

Charlene Richardson and Desmond Brown, two city workers, are responsible for organizing interventions and aid for young people flagged by the software.

"We put the picture together a bit more," said Ms. Richardson, who was recruited for the program after running youth centers in the area for two decades. "We know the computer doesn't always get it right."

Ms. Richardson and Mr. Brown came to the job with concerns that the system would unfairly target black boys. Now they are confident that the machine helps identify children who need help.

"This is not 'Minority Report,'" Mr. Brown said, referring to the 2002 Steven Spielberg movie. "We are not just taking the word of the computer."

The pair said they usually focused on the children with the highest risk scores, arranging home visits, speaking with their schools and finding mentors from the community.

"It's about seeing them as victims as well," she said.

'Does a computer know?'

Dr. Berk, the Penn professor who designed the algorithm used by the Philadelphia probation department, said controversy would fade as algorithms became more widely used.

He compared algorithms to the automatic pilot systems in commercial airliners. "Automatic pilot is an algorithm," he said. "We have learned that automatic pilot is reliable, more reliable than an individual human pilot. The same is going to happen here."

But people like Mr. Gates, whose future hangs in the balance, will take some convincing.

Sitting in the Philadelphia public defender's office, Mr. Gates said he was an easy person to read, pointing to the tattoos on his arms, which were meant to look like the bones under his skin. He understands machines. From a young age, he enjoyed dismantling computers and smartphones before putting them back together.

But Mr. Gates, whom we met through the defender's office, believed that a person could read him better than a machine.

"Does a computer know I might have to go to a doctor's appointment on Friday at 2 o'clock?" he asked.

Visiting the probation office so often can prevent him from getting the rest of his life on track. "How is it going to understand me as it is dictating everything that I have to do?" Mr. Gates asked.

Several weeks after his interview with The Times, he was allowed to make a short trip to Puerto Rico after a personal appeal to a judge. He always felt comfortable in front of his judge. The experience showed him the importance of a human touch.

"I can't explain my situation to a computer," Mr. Gates said. "But I can sit here and interact with you, and you can see my expressions and what I am going through."

Cade Metz reported from Philadelphia, and Adam Satariano from Rotterdam, the Netherlands, and Bristol, England.