

CASE STUDY

ALGORITHMIC DECISION- MAKING AND ACCOUNTABILITY

This is part of a set of materials developed by an interdisciplinary research team at Stanford University, led by Hilary Cohen, Rob Reich, Mehran Sahami, and Jeremy Weinstein. Their original use is for an undergraduate course on ethics, public policy, and technology, and they have been designed to prompt discussion about issues at the intersection of those fields. This case was written by Lauren Kirchner.

Stanford

ETHICS
TECHNOLOGY
&
PUBLIC POLICY

I. Introduction: A Call for Accountability in New York City

James Vacca, a popular Bronx-born Democrat, was serving his last term on the New York City Council when he decided to try to do something about a problem that had bothered him for a long time.

When the city had decided to shut down the fire department's ladder company on tiny City Island, Vacca was concerned: only one bridge connected the island to the mainland; wouldn't that pose a risk if fire trucks weren't able to get to a fire quickly enough, when traffic was high? The FDNY told him that they had a formula that determined where firefighters would be distributed throughout the city, and they couldn't do anything about its results.¹

Another time, when Vacca complained to the city that there weren't enough police officers assigned to several different precincts in his district, the NYPD told him that they had a formula that decided how police manpower would be allocated throughout New York. And when a parent in his district asked him why her child had been assigned to her sixth-choice public high school by the Department of Education's mysterious school-assigning algorithm, Vacca didn't have an answer for her.

Vacca had noticed for years that, just as many companies were using computer algorithms to make decisions -- from credit scoring to variable pricing to healthcare assessments -- cities were also increasingly relying on automated systems to make decisions about how to allocate their services. Algorithms were also assessing teacher performance in public schools, assigning people to public housing, and monitoring for fraud in government-assistance programs.² New York was at the forefront of this kind of technology.

It made sense, of course: more and more data about New Yorkers was being collected all the time, and it seemed like everything in the city government could run more efficiently if it could analyze that data in just the right way.

But, Vacca wondered, why did it have to be a secret? If the city was using software to make such important decisions, shouldn't the people on the receiving end of those decisions have the right to learn about them, and even to challenge them if they seemed wrong? And how could lawmakers like him advocate for their constituents, if they didn't even know what was going on?

In August 2017, Vacca proposed a bill in the Committee on Technology, of which he was the chairman.³ It mandated that city agencies put up on their websites the actual source code of all algorithms that they used to make their decisions, and to allow members of the public to "self-test" the algorithms by submitting their own data and getting the results. (See Exhibit 1 for the full text of proposed Int. No. 1696.) In a committee hearing about the bill later that year, Vacca said:

In our city it is not always clear when and why agencies deploy algorithms, and when they do, it is often unclear what assumptions they are based upon and what data they even consider... When government institutions utilize obscure algorithms, our principles of democratic accountability are undermined. As we advance into the twenty-first century, we must ensure our government is not "black-boxed," and I have proposed this legislation not to prevent city agencies from taking advantage of cutting edge tools, but to ensure that when they do, they remain accountable to the public.⁴

The idea that city agencies would publish the source code of all of their software online was unprecedented. Nothing like this bill had ever been passed anywhere in the country, and Vacca would soon find that passing it as-is would be an uphill battle.

II. “Black Box” Algorithms: What Could Go Wrong?

When he introduced this “algorithmic accountability” bill, as it came to be called, Vacca repeatedly stressed the need for transparency in government. But it wasn’t just about abstract ideals. He thought that algorithms deserved greater scrutiny because he had started to learn about ways that they could actually go very wrong.

Recent studies had begun to reveal troubling results that were affecting real people’s lives. It seemed that in some cases, the very systems that promised a corrective to human bias appeared to replace it with machine-based discrimination -- while presenting a veneer of mathematical accuracy and objectivity.

One example that Vacca cited as an inspiration for his algorithmic accountability bill was ProPublica’s 2016 investigation of COMPAS, a risk-assessment program and recidivism predictor that had begun to spread throughout the nation’s criminal justice system.⁵

COMPAS in Florida: Is it accurate? Is it fair?

In the 1970s, the jails of Broward County, Florida were so overcrowded that the county was forced to settle a lawsuit brought by its inmates over conditions there.⁶ This kicked off a long process of attempted improvements by the county. By 2008, the county sheriff’s office was still struggling with overcrowding. Instead of building more jails, they decided to try out a new method of parsing which arrestees would be sent there before trial, and which ones would be let out on bail.

Traditionally, bail decisions have been made by a judge, who made the determination on any number of factors: the crime that the person was accused of, his or her criminal history, record of making or missing court appearances, and so on. Under Broward’s new system, the pretrial hearing judge would still make that call, but he would have help from a “risk assessment” score. This score would show him see at a glance whether the county’s new software had deemed each prisoner a “high,” “medium,” or “low” risk of committing a new crime in the future if let out on bail, and if so, whether that crime would be violent.⁷

The ProPublica reporters wanted to learn more about the risk assessment software that Broward County was using, which was a program and called COMPAS (Correctional Offender Management Profiling for Alternative Sanctions), developed by Northpointe, Inc.⁸ So they sent public records requests for the formula itself under the Freedom of Information Act, but they were denied. Northpointe was a private company, and COMPAS was their product, so it had an interest in keeping those inner workings a secret. As a result, the algorithm was a black box.

ProPublica did get a copy of the 137-item questionnaire⁹ that each person had to fill out when they got arrested in Broward -- questions that ranged from the demographic to the psychological. (Sample questions were “Was one of your parents ever sent to jail or prison?” and “How often do you feel bored?”) The answers to these questions, along with the person’s criminal record and other information, apparently went into the COMPAS scores’ calculations. But there was no way to know how the different factors were weighted. So the reporters couldn’t exactly reverse-engineer the algorithm itself, but they could certainly analyze its outputs.

The county agreed to give ProPublica the risk scores assigned to 7,000 people who were arrested there during a two-year period. The reporters analyzed those scores alongside publicly available criminal records for those people, and jail information over the following years, which let them see how accurate the predictions turned out to be. In other words, did the people classified as being at a high risk of committing a new crime actually go on to commit one, and if so, was it a violent crime? Likewise, did supposedly low-risk people indeed steer clear of future criminal activity?

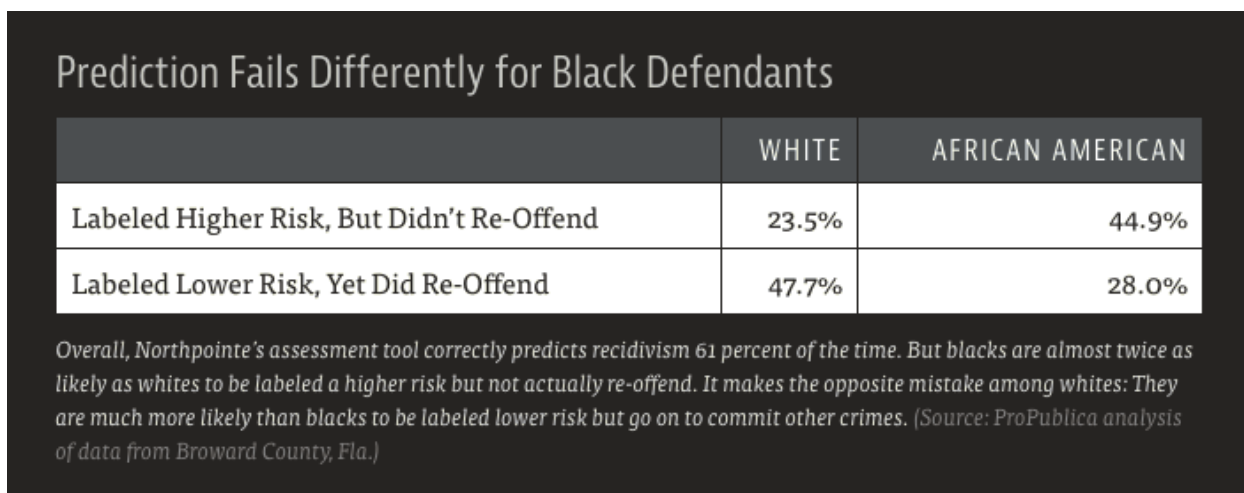
What they found was disturbing. Broward’s COMPAS scores appeared to be both inaccurate and racially biased.¹⁰ Here is a synopsis of their analysis:

The score proved remarkably unreliable in forecasting violent crime: Only 20 percent of the people predicted to commit violent crimes actually went on to do so.

When a full range of crimes were taken into account — including misdemeanors such as driving with an expired license — the algorithm was somewhat more accurate than a coin flip. Of those deemed likely to re-offend, 61 percent were arrested for any subsequent crimes within two years.

We also turned up significant racial disparities.... In forecasting who would re-offend, the algorithm made mistakes with black and white defendants at roughly the same rate but in very different ways.

The formula was particularly likely to falsely flag black defendants as future criminals, wrongly labeling them this way at almost twice the rate as white defendants. White defendants were mislabeled as low risk more often than black defendants.¹¹



So COMPAS assessments of white defendants and black defendants were both inaccurate at about the same rate, but the predictions for the two races were inaccurate in different directions. Black defendants’ scores tended to be false positives, while white defendants got false negatives. (See Exhibit 2 for excerpts from the ProPublica Machine Bias white paper, with more technical details.)

Nowhere in the 137-question surveys were defendants asked what race they were. So how could this be?

Algorithmic Bias

Programs like COMPAS, especially when they are employed in response to an existing problem (like overcrowded jails), can seem to promise “speed, efficiency and fairness” as well as financial savings.¹² Because they use computers and data, rather than arbitrary individual judgment, they are also often thought to be free of the bias and prejudice that can unconsciously affect human decisions. But as ProPublica’s analysis shows, technology is not necessarily neutral.

For instance, flaws or bad decisions in either the data that was used to train the program, or data used as inputs to the program’s process, can either intentionally or unintentionally lead to biased outcomes. Computer scientists discussing flawed or biased algorithms often use the phrase, “Garbage in, garbage out.” Or, as Solon Barocas and Andrew Selbst, researchers of the ethics and policies of artificial intelligence, explain,

an algorithm is only as good as the data it works with. Data is frequently imperfect in ways that allow these algorithms to inherit the prejudices of prior decision makers. In other cases, data may simply reflect the widespread biases that persist in society at large. In still others, [the algorithm] can discover surprisingly useful regularities that are really just preexisting patterns of exclusion and inequality.¹³

In 2016, The White House released a report on the promises and perils of Big Data, from credit scoring to hiring algorithms to college admissions programs.¹⁴ In each case, the report showed, algorithms present both opportunities and challenges; they can either mitigate or contribute to discriminatory results. The report stressed that “it is a mistake to assume [algorithms] are objective simply because they are data-driven.” The report identified four different ways that problems with the data algorithms use can potentially cause biased outcomes:

- 1) Poorly selected data
- 2) Incomplete, incorrect, or outdated data
- 3) Selection bias
- 4) Unintentional perpetuation and promotion of historical biases, “where a feedback loop causes bias in inputs or results of the past to replicate itself in the outputs of an algorithmic system.”¹⁵

The inner workings of the algorithm itself can also be a source of bias. An algorithm’s programmer must make decisions during the development process, about what will make it run efficiently, what conditions to optimize for, and so on. Depending on what assumptions and decisions go into the program’s design, flaws like these can also lead to bias on the back end:

- 5) Poorly designed matching systems
- 6) Personalization and recommendation services that narrow instead of expand user options
- 7) Decision-making systems that assume correlation necessarily implies causation
- 8) Data sets that lack information or disproportionately represent certain populations¹⁶

These problems can also be exacerbated in “machine learning” algorithms, which are designed to “learn” and evolve over time, in increasingly complex ways that their programmers can not necessarily anticipate, control, or even understand.¹⁷ The White House report argues that as complexity increases, “it may become more difficult to explain or account for the decisions machines make through this process unless mechanisms are built into their designs to ensure accountability.”¹⁸

COMPAS in Wisconsin: What about due process?

COMPAS, the risk assessment software used by the Broward County Sheriff's Office that ProPublica analyzed, is not a machine learning program; it's a simple set of scores, from 1 to 10, calculated from different so-called "risk factors" and other information.¹⁹

The main pre-trial hearing judge in Broward County told ProPublica that he relied on his own judgment, and not the COMPAS scores, to make decisions in his court. The makers of the software said that it wasn't meant to be used for higher-stakes decisions in court -- like, say, how long a sentence someone should get if they were found guilty of a crime. (The "Alternative Sanctions" at the end of the COMPAS acronym refers to different options the score might steer a judge to choose for a guilty defendant other than jail time -- like probation, "Drug Court," or mental health services.)

But at least at the time of the investigation, COMPAS scores were in fact being given to judges at the time of sentencing, in Arizona, Colorado, Delaware, Kentucky, Louisiana, Oklahoma, Virginia, Washington and Wisconsin.

Judges in these jurisdictions might, or might not, make reference to the score in court when announcing the bail determination. The defendant and defense attorney might or might not even be aware that the score exists, let alone how the COMPAS arrived at that score. For some criminal justice experts, this scenario raises concerns about due process, especially when the scores are affecting such vital decisions as how long someone will be sentenced to prison time.

"Risk assessments should be impermissible unless both parties get to see all the data that go into them," Christopher Slobogin, director of Vanderbilt Law School's criminal justice program told ProPublica. "It should be an open, full-court adversarial proceeding."²⁰

In fact, this very issue was the crux of an important criminal case that made it all the way to the Wisconsin Supreme Court.

In 2013, 34-year-old **Eric Loomis** pleaded guilty to driving a stolen car and evading police after he got caught driving a car that had been used in a shooting in LaCrosse County, Wisconsin. The LaCrosse circuit judge sentenced him to six years in prison. At Loomis's sentencing hearing, the judge mentioned a risk score that had apparently played into this decision. Judge Scott Horne told the court that Loomis had been "identified, through the COMPAS assessment, as an individual who is at high risk to the community."²¹ Loomis's prosecutors had also referred to his COMPAS score in their own arguments, saying: "the COMPAS report that was completed in this case does show the high risk and the high needs of the defendant. There's a high risk of violence, high risk of recidivism, high pre-trial risk; and so all of these are factors in determining the appropriate sentence."

Loomis had admitted guilt, but he objected to the way his sentence was being decided. In an appeal, Loomis argued that his right to due process had been violated, since this score was calculated through a private company's proprietary algorithm. He hadn't been given the right to challenge the evidence used against him in the sentencing hearing, he argued; he didn't have access to it at all.

The case made it all the way to the Wisconsin Supreme Court, which ruled on it in the summer of 2016. The justices ruled against Loomis's appeal, and said that it did not violate a defendant's due process rights when a risk assessment score was used in court. Loomis's sentence stood. Loomis then appealed to the U.S. Supreme Court, which declined to hear his case.²²

However, the Wisconsin court's opinion included an interesting caveat to the ruling. Judges needed to understand the limitations of these risk scores when they considered them, the opinion said, and the scores shouldn't be the "determinative factor" in the sentencing process. "Although we ultimately conclude that a COMPAS risk assessment can be used at sentencing, we do so by circumscribing its use," Justice Ann Walsh Bradley wrote.²³ To that end, the opinion recommended that a kind of warning label be attached to COMPAS scores in the pre-sentencing reports that judges receive, alerting the judge to certain "limitations and cautions" about the software. (See Exhibit 3 for the actual language that the court said should be appended.)

For its subtle analysis of risk assessments and their limitations, Christopher Slogobin called the decision "one of the most sophisticated judicial treatments of risk assessment instruments to date."²⁴

More broadly, cases like this raise an important question that extends beyond Wisconsin's jurisdiction, beyond the issue of risk assessments in sentencing, and even beyond the purview of the criminal justice system. Are citizens everywhere owed a certain measure of due process in the decisions that governments, non-governmental institutions, and companies make about them every day? What are people entitled to know about the inner workings of these decision-making systems, and what can remain inside a black box, in the name of security, or efficiency, or technological progress?

Accuracy vs. Fairness

In the *Loomis* opinion, Justice Bradley cited ProPublica's and others' analyses of COMPAS as indications that the software might suffer from racial bias, and that more research was needed. For instance, she wrote, "there is concern that risk assessment tools may disproportionately classify minority offenders as higher risk, often due to factors that may be outside their control, such as familial background and education."²⁵

Indeed, many of the 137 questions and answers in the COMPAS survey had to do with poverty and social marginalization.²⁶ The survey asked about gang activity in one's neighborhood, and whether or not one's house had a working telephone. Couldn't these be correlated with race? If so, this was unavoidable, said Tim Brennan, one of the founders of Northpointe and developers of COMPAS. Brennan told ProPublica that it would be hard to build a scoring system without questions like these. "If those are omitted from your risk assessment, accuracy goes down," he said.²⁷

For some computer science researchers, this raised some interesting questions about accuracy and fairness. Even if Brennan were correct, that racially-correlated data was necessary for accuracy, was that fair? Would it be possible in this scenario to have a risk-assessment scoring system to operate with both accuracy and fairness? Maybe not, as it turns out.

ProPublica's analysis of COMPAS had calculated that its rate of accuracy was the same for black defendants and white defendants: about 60 percent. But these predictions were accurate in opposite directions, so to speak. When the score was incorrect about black defendants, it tended to label them as being at a high risk of committing a new crime; when it was incorrect about white defendants, it labeled their risk as low.

Northpointe, the maker of the program, said that the program was fair and "racially neutral" because it had equivalent accuracy rates for both races. "The company said it had devised the algorithm to achieve this goal. A test that is correct in equal proportions for all groups cannot be biased, the company said."²⁸

So. How could a test be both fair and unfair at the same time?

Following the original COMPAS investigation, four groups of researchers working independently -- at Stanford University, Cornell University, Harvard University, Carnegie Mellon University, University of Chicago and Google -- all came to the same conclusion: Bias in risk assessment programs like this one is actually mathematically inevitable.²⁹

According to a ProPublica piece about this research:

The scholars set out to address this question: Since blacks are re-arrested more often than whites, is it possible to create a formula that is equally predictive for all races without disparities in who suffers the harm of incorrect predictions?

Working separately and using different methodologies, four groups of scholars all reached the same conclusion. It's not.

Revealing their preliminary findings on a Washington Post blog, a group of Stanford researchers wrote: 'It's actually impossible for a risk score to satisfy both fairness criteria at the same time.'³⁰ The problem, several said in interviews, arises from the characteristic that criminologists have used as the cornerstone for creating fair algorithms, which is that formula must generate equally accurate forecasts for all racial groups.

The researchers found that an algorithm crafted to achieve that goal, known as 'predictive parity,' inevitably leads to disparities in what sorts of people are incorrectly classified as high risk when two groups have different arrest rates.

The problem arose from the very beginning of the calculation. And the researchers couldn't find a way around it; it seemed to be an unresolvable Catch-22: "A risk score, they found, could either be equally predictive or equally wrong for all races — but not both."

One researcher who seemed to have the beginning of a solution was Alexandra Chouldechova, Assistant Professor of Statistics & Public Policy at Carnegie Mellon University. She found that, counter-intuitively, the only way for an equation like this to produce unbiased outcomes would be to weight factors differently for people of different races on the input side.³¹ But is that fair?

And there are other Catch-22s, as well: another group of researchers used the same COMPAS dataset to explore how the risk assessment system could optimize for "improving public safety," or for "satisfying prevailing notions of algorithmic fairness," but not both.³²

The puzzle that COMPAS presented continues to intrigue computer scientists and researchers working on how to improve fairness in machine learning systems.

For ethicists, or for anyone thinking about their expectations of justice and transparency in the institutions they interact with, "fairness" is an abstract but simple concept. But for computer scientists, there are many competing ideas about what fairness means, in the technical or mathematical sense. Every programmatic system is embedded with different assumptions about what methods are most efficient, and what outcomes are the most desirable. These assumptions and choices will also necessarily have different trade-offs.

One classic "trade-off" example that has long been discussed in philosophy classes is the "trolley car" thought experiment. (Briefly: is it more ethical to do nothing and let a runaway trolley car kill five people, or to pull a lever and cause the car to switch tracks and kill only one person?) But in computer science, there are trolley-car problems at every turn. In fact, **Arvind Narayanan**, who teaches computer science at Princeton, gave a talk in February 2018 summarizing 21 definitions of fairness.³³

Narayanan has argued that the search for “one true definition” of fairness is “a wild goose chase,” and that it is neither possible nor necessary for computer scientists to all agree on one rubric.³⁴ This might also seem counter-intuitive to those outside the field. As he explained in a series of tweets in November 2017:

It’s true that in most subfields of [computer science], there are usually a small/manageable number of objective desiderata. In machine learning, there are standard benchmarks for most problems. Say for object detection, measure accuracy on ImageNet. It doesn’t that much matter which standard we pick, just that we have a standard. In other fields, definitions matter more, but the community can agree on a definition, such as semantic security in cryptography. Well, this is not going to happen for fairness in [machine learning]. Fairness is an inherently, deeply social concept. Depending on the context, it will translate into different subsets of mathematical criteria. And all the math in the world won’t fully capture what we mean by fairness. [...] This calls for new structures of collaboration, which is what makes the field so exciting.³⁵

Man vs. Machine

Proponents of finding machine-learning solutions to complex social problems will often say, so what if an algorithm isn’t quite perfect? It must be better than the old way of doing things, where judges don’t have any scores to help them make decisions, right?

Judges are flawed human beings. They are necessarily affected by mood swings, exhaustion, and their own individual (conscious and unconscious) biases. Despite their mandate to approach the facts of each case before him or her with an open mind and a clean slate, judges aren’t calculators. They are often under pressure to make decisions quickly and intuitively.

Judge Jerome Frank once said, “justice is what the judge ate for breakfast.” In fact, a study in Israel found that judges were twice as likely to grant a defendant parole at the beginning of a court session, right after they would have eaten a meal or a snack, than at the end of a session, when they were presumably hungry and tired.³⁶ Other research has shown how race³⁷, gender³⁸ and age³⁹ can all contribute to disparities in sentencing⁴⁰ -- not to mention the vast disparities between the treatment of defendants from one jurisdiction to another.⁴¹

So, given that judges left to their own devices can make biased (or at the very least arbitrary) decisions, and criminal-risk-scoring algorithms apparently have their own problems, how do they compare to each other? Do the scores help the judges make decisions in court that are more, or less, biased?

This turns out to be another unanswerable question, at least for now.

Also following ProPublica’s investigation of COMPAS, **Rose Eveleth** addressed that question in a piece for Vice magazine’s Motherboard blog.⁴² Her headline says it all: “Does Crime-Predicting Software Bias Judges? Unfortunately, There’s No Data.”

“It’s important to question biases of all kinds, both human and algorithmic, but it’s also important to question them in relation to one another,” wrote Eveleth. “And nobody has done that.” Eveleth had reached out to both Northpointe (the maker of COMPAS) and the New York Division of Criminal Services (which had approved the use of COMPAS in that state), and both had told her that they were not aware of any research comparing sentencing outcomes with and without COMPAS scores. They could point to studies of the scores’ accuracy in predicting outcomes, but nothing that compared the

outcomes of judges who used COMPAS and those who didn't (or even the same judges, before and after COMPAS). Eveleth writes:

This shouldn't be hard to find out: ideally you would divide judges in a single county in half, and give one half access to a scoring system, and have the other half carry on as usual. If you don't want to A/B test within a county—and there are some questions about whether that's an ethical thing to do—then simply compare two counties with similar crime rates, in which one county uses rating systems and the other doesn't. In either case, it's essential to test whether these algorithmic recidivism scores exacerbate, reduce, or otherwise change existing bias. [...]

As far as I can find, and according to everybody I've talked to in the field, nobody has done this work, or anything like it. These scores are being used by judges to help them sentence defendants and nobody knows whether the scores exacerbate existing racial bias or not.⁴³

Research like this may be more complicated than she thinks, said Suresh Venkatasubramanian, a computer scientist who is prominent in the growing field of fairness in machine learning. But he agreed that it was “bewildering” that no one seemed to be trying to do it at all. “I often wonder if it's because part of the rationale for using automated methods is ease of use and speed, and having to do elaborate studies on their efficacy defeats the purpose,” Venkatasubramanian told Eveleth.⁴⁴

III. The Debate in NYC

On the day the New York City Council Committee on Technology held a hearing about Vacca's proposed algorithmic bill in October 2017, the committee room at City Hall was packed. Tech bloggers and news reporters, city employees and nonprofit staffers filled every seat and leaned against the walls in the back and sides of the chamber.

Vacca's proposal was, as far as everyone could tell, the first of its kind in any part of the U.S., so curiosity was high. Could a bill like this actually pass? Should it? As he banged the gavel to start the hearing, Vacca remarked that it was the best attendance that the Tech Committee had ever had.

The committee invited a wide range of stakeholders to share their thoughts about the bill. But first, Vacca set out the crux of the problem in his introductory remarks. “While it is undeniable that these tools help city agencies operate more effectively and do offer residents more targeted, impactful services, algorithms are not without issue,” he said. “These tools seem to offer objectivity, but we must be cognizant of the fact that algorithms are simply a way of encoding assumptions; that their design can be biased, and that the very data they possess can be flawed.”⁴⁵

Many people who spoke at the hearing who were wholeheartedly in favor of its wide-sweeping mandate for transparency, including the publication of city agencies' algorithms' source code. Others only liked some parts of the bill but had concerns with the rest of it. Still others seemed to think the bill, as proposed, it was either wildly impractical, or dangerous, or both.

Calls for transparency

Advocates from civil rights groups and criminal-justice nonprofits were among the most enthusiastic about the bill. Attorneys from the Legal Aid Society, which had been fighting for years to gain access to the source code of what its DNA experts believed was a flawed crime-scene-DNA analysis

program,⁴⁶ spoke about the need for transparency of a wide range of programs that affected New Yorkers' lives -- including bail determinations, family court hearings, juvenile delinquency proceedings, parole hearings, and sex offender registration.

One tool in particular, which was starting to be used to help judges predict which defendants would fail to appear back in court if they were let out on bail, "to our knowledge has never been independently studied or verified, and anonymized data and source code has never been released to independent third parties," said Joshua North, a Legal Aid staff attorney.

Staffers from the Bronx Defenders and the Brooklyn Defenders concurred. "[T]his tool...would be used by judges in thousands of cases across the city, tens of thousands of cases every year, in making bail determinations," said Scott Levy from the Bronx Defenders. "That is determining whether a New Yorker returns to their family and community after they are arraigned in criminal court, or whether they spend days, months, or even years sitting on Rikers Island."

For years, New York politicians had been talking about closing Rikers Island;⁴⁷ Levy said that he and his colleagues believed that these types of algorithms could actually lead to an increase in pre-trial detention there, by giving judges the false promise of accuracy.

Rachel Levinson-Waldman, Senior Counsel to the Liberty and National Security Program at the Brennan Center for Justice, also spoke about New York's predictive policing program, which used crime statistics to train algorithms that could supposedly predict where crimes might occur in the future -- and who might commit them. Software told The New York Police Department how many officers to deploy to patrol certain neighborhoods, and at what times. Publicly available documents suggested the NYPD expected to spend about \$45 million on these programs in the next five years, she said.

The problem, Levinson-Waldman believed, was that these algorithms could exacerbate racial bias in policing, by sending more officers to neighborhoods that were already over-policed to begin with. Training data might show concentrated crime in certain places where police had been more likely to stop people in the past, not where people were more likely to be committing crime in the future. "In addition, there is little hard proof that predictive policing is actually effective in predicting and reducing crime," said Levinson-Waldman. "One phrase often used is that predictive policing predicts policing. It does not predict crime."

"This is like minority report," City Council Member David Greenfield remarked.

Fears about security risks

Levinson-Waldman also said that the Brennan Center had submitted Freedom of Information Act requests to the NYPD about its software, including the programs' source code. But the NYPD had denied the request, arguing that if that code were to be published, potential criminals could learn where police officers would be at certain times (and where they wouldn't be), and then use that information to commit crimes without getting caught.

Like the NYPD, two representatives from the Mayor's office who spoke at the hearing said that there would be dire security risks to revealing the inner workings of the city's decision-making algorithms.

"It is the opinion of our cybersecurity experts that publishing algorithms would generate considerable risk, providing a roadmap for bad actors to attack crucial city systems," said Don Sunderland, a deputy commissioner at the city's Department of Information Technology. He also

argued that opening up the software for testing by New Yorkers was a bad idea because it “could empower users to fabricate answers that will get them the response they want.”

Craig Campbell, Special Adviser to the Mayor’s Office on Data Analytics, was at the hearing to answer questions about data analytics and open data projects in the city government. Vacca asked Campbell about the formula that the government used to distribute firefighters throughout the city, and why even he, as a lawmaker, wasn’t able to know more about how it worked. But Campbell said he was unfamiliar with it, that he hadn’t worked with the FDNY at all. He said his department provided data-analytics assistance to the city’s agencies as they needed it, but it didn’t oversee what all of the agencies did.

Vacca seemed annoyed. “Is there no centralized oversight over when agencies deploy potentially complex data analytics?” he asked.

“There is not, and I would argue that it’s probably better that way,” Campbell replied. “The city has been organized with the idea of putting the technology as closely as possible to the actual operational functionality that the agencies have to deliver.”

The open-source question, and citizens’ privacy

Vacca had another question. Campbell and Sunderland had said that publishing the city’s algorithms’ source code would present security risks, but Vacca said he had heard from computer science experts that open-source software could actually be *more* secure. So which one was it?

Campbell explained that software that was developed in an open-source, collaborative way was one thing, but that the process of making public once-proprietary software was different. The city’s algorithms hadn’t designed from the beginning to protect sensitive information from people wanting to do harm, he said. Reverse-engineering that would be too difficult.

Another broad objection that Sunderland raised in his testimony was that the proposed bill would put the information of the most vulnerable New Yorkers -- those who are disabled or receive government benefits, for instance -- at risk of being exposed to people or institutions who might discriminate against them because of it, or to people who might want to take advantage of that information in other ways. Taline Sanassarian, the policy director for the technology industry trade group Tech NYC, agreed with those privacy concerns. “Indeed, one may look no further than the recent breaches of data, including Equifax, which affected as many as 145 million Americans,” she said, “in which sensitive personnel information was stolen from current and former government employees and contractors.”⁴⁸

City Council Member Greenfield posed that same question to some of the criminal defense advocates during the hearing: *What about* those security risks? Should the city have concerns about the confidentiality of New Yorkers’ data, if the data were to be open to the public?

The Bronx Defenders’ Scott Levy said that no, he didn’t think so. Releasing the training data that had been used in the past to train predictive policing models, for instance, wouldn’t put any individuals at risk because it would all be “anonymized, and randomized, and essentially ‘clean.’” Yung Mi Lee from the Brooklyn Defenders took a stronger stance, arguing, “when we’re talking about constitutional protections versus possible security risks that aren’t even realized and may never happen, I think our constitutional protections have to take precedence.”

How would it work?

Aside from advocates and city employees, there were also technologists on hand. Their reviews of Vacca's proposed bill were mixed.

Noel Hidalgo, the Executive Director of the nonprofit group of data scientists and hackers Beta NYC, spoke first. He said he thought the bill was an important step towards open and equitable government. "[T]o be perfectly blunt, our future of democracy is at stake," said Hidalgo. "If we refuse to hold algorithms and their authors accountable, we outsource our government to the unknown."

But how would that accountability actually work, others asked? Even if it's taken as a given that people should have access to the formulas that are making decisions about them, what would that actually look like?

One computer science researcher who has done testing of complex algorithms, Julia Powles from Cornell Tech, said, "It often takes me thousands of queries, depending on the context, to be able to do the necessary third-party testing in the public interest of algorithmic systems." Would individual New Yorkers really be able to test the formulas themselves?

Charlie Moffett, a graduate student at the NYC Center for Urban Science and Progress, who had researched civic algorithms, pointed out that source code is too complex for most people to understand. He thought it was more important to make the *outcomes* of algorithms accessible, rather than the algorithms' design. If algorithms are to be held to greater standards of transparency, he argued, then that burden should fall on the designers and the vendors, not on New Yorkers. And what the city's agencies should be doing, he added, was using their leverage over the data analytics firms competing for their work, in order to build more transparency requirements into the contracts in the first place.

What next?

After the hearing, a reporter asked Vacca what he thought about the day's testimonies, and what he thought the chances were of the bill passing after listening to what everyone had to say. "I don't think the bill's going to pass as it is," Vacca said. "I think we're going to have to make some changes, because it is complex, and it is detailed."⁴⁹

Vacca said he thought many of the witnesses had brought up great points, especially the importance of keeping ongoing police investigations confidential, and the question about how to keep third-party vendors to the same standards of transparency as the city agencies themselves.

Despite the challenges ahead, Vacca added that he was intent on passing the bill in some form before the end of his term, just two and a half months away. He had served on the City Council since 2006, and term-limit rules prevented him from running again. So, he really wanted something to leave behind for his colleagues on the committee to work with.

IV. Food for Thought / Looking to the Future

What does regulation, or legal recourse, look like?

It's true that New York's algorithmic accountability bill was the first of its kind in the country,⁵⁰ but there are precedents that could serve as guides for future efforts to regulate. In the 1970s, the US passed the Fair Credit Reporting Act (FCRA) and the Equal Credit Opportunity Act (ECOA). These were established to help protect consumers from being harmed by a very opaque and very significant algorithm, the credit score. The FCRA was born out of the idea that people should have the right to know what kinds of information were going into the score, and further, have the right to challenge those inputs. It's a more focused version of what Vacca was after with his bill.

Overseas, the European Union has made headway in expanding this principle to commercial algorithms writ large. The General Data Protection Regulation (GDPR) will go into effect in May 2018 and will apply to *all* companies and people in the EU.⁵¹ The GDPR stipulates that all algorithmic decisions be accountable to the individuals they affect -- that citizens have the "right to explanation" about how their personal data has been used in that decision-making process.⁵² And because of the increasingly global nature of digital companies, the impact of this philosophy will have ripples outside of the EU as well.

Academic thinkers working in the field of algorithmic accountability have other ideas about how regulation could potentially work in the future. Some have proposed the creation of formal regulatory processes within government institutions, like the Federal Trade Commission.⁵³

Another intriguing idea is to apply "disparate impact" theory to algorithmic systems within the context of the US legal system.⁵⁴

First applied to employment law in a 1971 Supreme Court case,⁵⁵ disparate impact is the theory that systems can be held responsible for discriminatory *results* even if there was no discriminatory *intent*. In 2015, another landmark Supreme Court decision applied the theory to housing discrimination as well.⁵⁶ Just as the analysis of COMPAS risk assessment scores focused on the outcomes of the software rather than its proprietary design, perhaps legal challenges to other algorithms could work the same way.

For instance, a Bloomberg investigation of Amazon found a racial disparity in the company's delivery zones; areas of cities that had access to Amazon's same-day delivery service were predominantly white, while predominantly black areas were more likely to be excluded.⁵⁷ Similarly, ProPublica found that Princeton Review was charging different prices to customers in different ZIP codes for the same online test-prep service -- and a curious side effect of this was that Asians were twice as likely to be shown a higher price than non-Asians.⁵⁸

No one thinks that Amazon or Princeton Review intentionally set out to treat its customers differently based on their race. These were likely the unintentional results of automated geographic market research and variable pricing models. But under disparate impact theory (as well as in journalism), the results are what matter most.

However, this type of legal remedy for unintentionally discriminatory algorithms remains just an academic idea for now. Solon Barocas and Andrew Selbst, who posed this thought experiment in their *California Law Review* article "Big Data's Disparate Impact," wrote that expanding disparate impact theory to algorithms in court "will be difficult technically, difficult legally, and difficult politically."⁵⁹

What does access look like?

Compared to litigating the harm done to a person by algorithm, a somewhat more straightforward (though still contentious) legal question is whether or not a person has the right of access to a proprietary algorithm's inner workings. This was the central question in the *Loomis* case in the Wisconsin Supreme Court. This debate also played out in California's appeals court when the American Civil Liberties Union and the Electronic Frontier Foundation intervened to fight for the right of a man accused of serial rape to have access to the source code of a commercially-developed forensic DNA analysis program called TrueAllele.⁶⁰

And after the conclusion of a gun possession case in New York, ProPublica filed a motion arguing that the source code of another proprietary DNA analysis program should be lifted.⁶¹ A federal judge unsealed the code for the Forensic Statistical Tool, a program that had been invented by scientists working for the city in the Office of the Chief Medical Examiner but whose development had been surrounded by secrecy and controversy for years.⁶² When ProPublica obtained the code, they published it online and invited coders to evaluate it.⁶³

News organizations and advocacy groups have also relied on the Freedom of Information Act to attain similar goals. The Electronic Privacy Information Center and MuckRock both have many ongoing projects that involve canvassing states and cities across the country with FOIA requests, trying to gather information about the algorithms used in criminal justice proceedings, for instance.⁶⁴

Fighting in the courts or fighting with FOIAs for source code access are lengthy and piecemeal projects, however. And is source code really the key to algorithmic accountability, after all?

Several witnesses in the New York City Council hearing had argued that source code alone wouldn't tell average New Yorkers that much at all - nor would "self-testing" by the public be as simple as it might sound. (Case in point: see Exhibit 4 for a sample of the FST source code that ProPublica obtained and published.)

Other researchers have agreed that the source code issue is a bit of a red herring. Writing about the *Loomis* decision, former Facebook engineer **Ellora Israni** explained:

The fixation on examining source code reflects a fundamental misunderstanding about how exactly one might determine whether an algorithm is biased. It is highly dubitable that any programmer would write 'unconstitutional code' by, for example, explicitly using the defendant's race to predict recidivism anywhere in the code. Examining the code is unlikely to reveal any explicit discrimination. ... The answer probably lies in the inputs given to the algorithm. Thus, the defendant should be asking to see the data used to train the algorithm and the weights assigned to each input factor, not the source code.⁶⁵

So aside from raw code, other options for disclosures about decision-making algorithms could include the original sets of training data and the weighted factors that go into the algorithms' design. Or, to make things even more accessible to the non-programming public, governments could provide visual charts breaking down a scoring system into discrete decision points, or even just written explanations of the basic logical assumptions behind the programs' conception.

Whatever the means of access, put simply, the spirit of algorithmic accountability as a broader ideal is public transparency. It's the ability of the people to be aware of, and participate in, the decisions that automated systems increasingly make about their lives.

That is why algorithms that get developed out in the open from the very start garner such praise -- like Allegheny County, Pennsylvania's predictive analytic system for child welfare cases.⁶⁶ Time will tell whether caseworkers armed with risk scores will do a better or worse job at protecting children from abuse than they did without, but county officials say they are doing this "the right way": slowly, and out in the open. According to a *New York Times Magazine* piece about the Allegheny Family Screening Tool:

It is owned by the county. Its workings are public. Its criteria are described in academic publications and picked apart by local officials. At public meetings held in downtown Pittsburgh before the system's adoption, lawyers, child advocates, parents and even former foster children asked hard questions not only of the academics but also of the county administrators who invited them.⁶⁷

Child and family advocates, including the ACLU of Pennsylvania, told the Times reporter that they approved of what they saw so far.

What does responsibility look like?

The first step in the implementation of any algorithm, public or private, is its design. The White House's 2016 report says that we should make sure that Big Data doesn't exacerbate society's systematic disadvantaging of certain groups, and the way to do this is with "equal opportunity by design"—designing data systems that promote fairness and safeguard against discrimination from the first step of the engineering process and continuing throughout their lifespan."⁶⁸

Thoughtful design is even more vital for machine-learning or neural-network programs, which become more and more complex and less and less scrutable to its own designers (not to mention less susceptible to external audits) over time.

So what responsibility do the computer scientists, machine-learning developers, coders and researchers themselves have, to make sure their designs do more good than harm? Do they have any? If so, what guidelines should they be expected to follow?

This has been an ongoing conversation in the tech community for several years. At the center of the conversation is a growing group of researchers and practitioners called Fairness, Accountability and Transparency in Machine Learning, or FAT/ML for short. What began as a small conference workshop to explore ethical issues in computer science has grown to a large international community that gathers at standing-room-only annual summits.

Given the "significant social impact" of so many decision-making algorithms, FAT/ML's leadership wrote a set of guiding principles that are now published on the group's website.⁶⁹ The guidelines include Responsibility, Explainability, Accuracy, Auditability and Fairness. The site also includes a long list of questions and suggestions that an algorithm's designers should consider throughout the process, such as, "Are there particular groups which may be advantaged or disadvantaged, in the context in which you are deploying, by the algorithm/system you are building?" and "Talk to people who are familiar with the subtle social context in which you are deploying". (See Exhibit 5 for the full text of FAT/ML's "Principles for Accountable Algorithms and a Social Impact Statement for Algorithms.")

All of FAT/ML's guiding principles follow from a single premise, which is:

Algorithms and the data that drive them are designed and created by people -- There is always a human ultimately responsible for decisions made or informed by an algorithm. 'The algorithm did it' is not an acceptable excuse if algorithmic systems make mistakes or have undesired consequences, including from machine-learning processes.⁷⁰

The Association for Computing Machinery, a group of over 100,000 computer science practitioners and educators around the world, has also developed its own set of "Principles for Algorithmic Transparency and Accountability."⁷¹ They fall under the categories of Awareness, Access and Redress, Accountability, Explanation, Data Provenance, Auditability, and Validation and Testing. And a research center called "The AI Now Institute" published its own set of recommendations for both responsible design and responsible deployment by companies and governments.⁷² (See Exhibits 6 and 7 for more on these.) The AI Now Institute also put out a step-by-step guide for creating "Algorithmic Impact Assessments."⁷³ If the policymakers of a new law, or the builders of a new construction project, are expected to systematically consider the impact on the people and the environments that will be affected by their choices, why aren't the designers and the customers of complex algorithms expected to do the same?

Finally, when computer scientists and machine-learning developers consider the most ethical way to build a program, it's possible that the most ethical choice would be to choose to not build it at all. When AI Now's co-founder, Kate Crawford, spoke during the closing panel of FAT/ML's 2016 conference, she suggested that everyone in the tech community should consider his or her personal politics and ethical beliefs when deciding what programming jobs to accept.⁷⁴ "What will we do with these tools, and what won't we do?" she asked.

As we know, machine learning is already being used in predictive policing, in the criminal justice system, and in tracking refugee populations here and in Europe. Might it also be used to track undocumented workers? Might it be used to help create the 'Muslim registry' that we've heard so much about? I would like to say this to this community here, but also to Silicon Valley more generally, that I think we all have a responsibility right now to decide, where we will put our work, and where we will say, this type of work is not ethical.⁷⁵

Despite these ongoing and passionate (if abstract) conversations, it's clear that many program developers are still grappling with these questions in their day-to-day work. According to the 2018 Stack Overflow Survey of over 100,000 developers:

Only tiny fractions of developers say that they would write unethical code or that they have no obligation to consider the ethical implications of code, but beyond that, respondents see a lot of ethical gray. Developers are not sure how they would report ethical problems, and have differing ideas about who ultimately is responsible for unethical code.⁷⁶ (See Exhibit 8 for some examples of this survey's results.)

V. Conclusion: The Fate of New York's Bill

James Vacca's suspicions that he expressed at the end of the October hearing proved to be correct. The wide-sweeping bill that he had originally proposed -- mandating the publication of the source codes of all of the algorithms used by all of the city's agencies, and allowing New Yorkers to test them online -- was bound to fail. But in the end, he and his staff took everyone's objections into consideration, and presented a new, watered-down version to the City Council in December.

The revised bill wouldn't require the city agencies to publish anything online themselves. It would establish a task force that would examine exactly what decision-making algorithms are in use throughout the city government, and whether any of them appear to discriminate against people based on age, race, religion, gender, sexual orientation or citizenship status. It would also explore ways for New Yorkers to better understand these algorithms, and to challenge any decisions that they didn't agree with.

The task force would be made up of a wide range of experts and advocates, including "persons with expertise in the areas of fairness, accountability and transparency relating to automated decision systems." And then it would deliver a report and recommendations to the city council within a year and a half. (See Exhibit 9 for the full text of the final bill, Int. No. 1696.)

This new bill passed unanimously.⁷⁷

Not everyone was happy about the changes that the bill had to go through in order to get passed, however. A caveat had been added to the task force's mandate, stating that agencies would not be required to hand over any information to the task force that would

"interfere with a law enforcement investigation or operations, compromise public health or safety, or that would result in the disclosure of proprietary information."

In other words, participation in the task force's project was voluntary. Some civil rights and criminal defense advocates who had enthusiastically supported the original bill wondered, what's to stop law enforcement from making a blanket denial to the task force on these grounds?

Julia Powles, who had testified at the hearing, called it a "bold" but "flawed" bill in an opinion piece for *The New Yorker*.⁷⁸ "[N]ow the task force will have to rely on voluntary disclosures as it studies how automated systems are designed, procured, and audited," she wrote. "For a government body without real legal powers, this will be a Herculean, or perhaps Sisyphean, undertaking."

Powles also quoted Ellen Goodman, a Rutgers Law School professor, who was disappointed that the bill didn't include anything about how the city government could put pressure on third-party vendors to make their proprietary programs more transparent to lawmakers and the public. "For many of these vendors, it's the biggest customer they'll get," said Goodman. "If New York doesn't use that power to make systems accountable, who will?"⁷⁹

Rashida Richardson, Legislative Counsel for the New York Civil Liberties Union, said that in addition to her concerns about the bill's carve-outs, she also wondered about what the composition of the task force would be.⁸⁰ The bill gives the power to the Mayor to appoint whomever he wanted, and Richardson said that he shouldn't choose all government workers, or all technologists, or all academic researchers. More importantly, though, she wondered what the ultimate outcome of this task force report would be.

“The task force could be perfect, they could have access to all the information they need to make the perfect recommendations, but then ultimately if the city doesn’t adopt any of them, then that’s the ultimate concern,” said Richardson.

When asked whether the task force would have the access and the authority it needed, given these caveats, Vacca said that he didn’t anticipate problems.⁸¹ The city council would have oversight over the process the whole time, he explained, and if the task force was having trouble getting information, the city council members could step in and correct that.

As for the “self-testing” idea that he had originally wanted to include, Vacca explained that this bill would accomplish something similar, by exploring ways to make the formulas generally accessible and understandable to the public. So, for instance, agency websites should disclose what kinds of software they are using and list the factors that go into them. And these should be written out in easy-to-read explanations, or illustrated with charts or decision trees -- not replicated in computer code.

Overall, Vacca was happy with the bill and proud that he could leave it behind as part of his legacy when he left office. He said that he hoped that other cities and states would take note of this first effort and start to have similar conversations about the algorithms being used by their government agencies. He also said he knew that there were big changes to the bill, from what he had originally proposed to what he ended up with. But he was glad that he could leave something behind for his colleagues to work with.

“You know, it’s part of the legislative process, that, when you have a bill, the bill is negotiated,” said Vacca. “I thought I would go for the whole thing, and then see what I got. And I think I got something good.”

Exhibit 1 - Full text of New York City’s “algorithmic accountability” bill, as proposed in August 2017

Int. No. 1696

By Council Member Vacca

A Local Law to amend the administrative code of the city of New York, in relation to automated processing of data for the purposes of targeting services, penalties, or policing to persons.

Be it enacted by the Council as follows:

Section 1. Section 23-502 of the administrative code of the city of New York is amended to add a new subdivision g to read as follows:

g. Each agency that uses, for the purposes of targeting services to persons, imposing penalties upon persons or policing, an algorithm or any other method of automated processing system of data shall:

- 1. Publish on such agency’s website, the source code of such system; and
- 2. Permit a user to (i) submit data into such system for self-testing and (ii) receive the results of having such data processed by such system.

§ 2. This local law takes effect 120 days after it becomes law.

MAJ

LS# 10948

8/16/17 2:13 PM

Exhibit 2 - An excerpt from the white paper explaining the methodology of ProPublica’s “Machine Bias” investigation

[W]e investigated whether certain types of errors – false positives and false negatives – were unevenly distributed among races. We used contingency tables to determine those relative rates

	All Defendants		Black Defendants		White Defendants		1		
	Low	High	Low	High	Low	High			
Survived	2681	1282	Survived	990	805	Survived	1139	349	e logistic s in the ving tables
Recidivated	1216	2035	Recidivated	532	1369	Recidivated	461	505	
FP rate: 32.35			FP rate: 44.85			FP rate: 23.45			
FN rate: 37.40			FN rate: 27.99			FN rate: 47.72			
PPV: 0.61			PPV: 0.63			PPV: 0.59			
NPV: 0.69			NPV: 0.65			NPV: 0.71			
LR+: 1.94			LR+: 1.61			LR+: 2.23			
LR-: 0.55			LR-: 0.51			LR-: 0.62			

These contingency tables reveal that the algorithm is more likely to misclassify a black defendant as higher risk than a white defendant. Black defendants who do not recidivate were nearly twice as likely to be classified by COMPAS as higher risk compared to their white counterparts (45 percent vs. 23 percent). However, black defendants who scored higher did recidivate slightly more often than white defendants (63 percent vs. 59 percent).

The test tended to make the opposite mistake with whites, meaning that it was more likely to wrongly predict that white people would not commit additional crimes if released compared to black defendants. COMPAS under-classified white reoffenders as low risk 70.5 percent more often than black reoffenders (48 percent vs. 28 percent). The likelihood ratio for white defendants was slightly higher 2.23 than for black defendants 1.61.

Exhibit 3 - An excerpt from the Wisconsin Supreme Court's opinion on *Wisconsin v. Loomis*

¶98 Thus, a sentencing court may consider a COMPAS risk assessment at sentencing subject to the following limitations. As recognized by the Department of Corrections, the PSI instructs that risk scores may not be used: (1) to determine whether an offender is incarcerated; or (2) to determine the severity of the sentence. Additionally, risk scores may not be used as the determinative factor in deciding whether an offender can be supervised safely and effectively in the community.

¶99 Importantly, a circuit court must explain the factors in addition to a COMPAS risk assessment that independently support the sentence imposed. A COMPAS risk assessment is only one of many factors that may be considered and weighed at sentencing.

¶100 Any Presentence Investigation Report (“PSI”) containing a COMPAS risk assessment filed with the court must contain a written advisement listing the limitations. Additionally, this written advisement should inform sentencing courts of the following cautions as discussed throughout this opinion:

- “The proprietary nature of COMPAS has been invoked to prevent disclosure of information relating to how factors are weighed or how risk scores are determined
- Because COMPAS risk assessment scores are based on group data, they are able to identify groups of high- risk offenders – not a particular high-risk individual.
- Some studies of COMPAS risk assessment scores have raised questions about whether they disproportionately classify minority offenders as having a higher risk of Recidivism.
- A COMPAS risk assessment compares defendants to a national sample, but no cross-validation study for a Wisconsin population has yet been completed. Risk assessment tools must be constantly monitored and re-normed for accuracy due to changing populations and subpopulations.
- COMPAS was not developed for use at sentencing, but was intended for use by the Department of Corrections in making determinations regarding treatment, supervision, and parole.”

Exhibit 4 - A sample of the Forensic Statistical Tool's source code

```
/// <summary>
/// This function loops according to the race calculating the frequency, the numerator (Pn), the denominator (Pd)
/// and lastly the Ratio (P) for each race.
/// </summary>
/// <param name="Theta"></param>
/// <param name="compareMethod"></param>
/// <param name="intProfileNo"></param>
/// <returns></returns>
public Dictionary<string, float> DoCompare(
    DataTable dtFrequencies = null,
    Dictionary<string, Dictionary<string, double>> perRaceLocusDenominatorCache = null,
    Dictionary<string, Dictionary<string, float>>> perRaceLocusFrequencyCache = null,
    Dictionary<int, Dictionary<string, Dictionary<string, float>>> perReplicateDropOutCache = null,
    Dictionary<string, Dictionary<int, List<AllelesPair>>> numeratorPermutationCache = null,
    Dictionary<string, Dictionary<int, List<AllelesPair>>> denominatorPermutationCache = null)
{
    // set up our result dictionary, database connection, and race table
    DNAComparisonResult = new Dictionary<string, float>();
    myDb = myDb ?? new Database();
    raceTable = raceTable ?? myDb.getAllEthnics();

    int compareMethod = comparisonData.CompareMethodID;
    // check to see if we should remove any loci based on the fact that the frequencies for the alleles at that locus
    CheckFrequencyForRemoval(dtFrequencies);
}
```

Exhibit 5 - The full text of FAT/ML's "Principles for Accountable Algorithms and a Social Impact Statement for Algorithms."

Principles for Accountable Algorithms

Automated decision making algorithms are now used throughout industry and government, underpinning many processes from dynamic pricing to employment practices to criminal sentencing. Given that such algorithmically informed decisions have the potential for significant societal impact, the goal of this document is to help developers and product managers design and implement algorithmic systems in publicly accountable ways. Accountability in this context includes an obligation to report, explain, or justify algorithmic decision-making as well as mitigate any negative social impacts or potential harms.

We begin by outlining five equally important guiding principles that follow from this premise:

Algorithms and the data that drive them are designed and created by people -- There is always a human ultimately responsible for decisions made or informed by an algorithm. "The algorithm did it" is not an acceptable excuse if algorithmic systems make mistakes or have undesired consequences, including from machine-learning processes.

Responsibility

Make available externally visible avenues of redress for adverse individual or societal effects of an algorithmic decision system, and designate an internal role for the person who is responsible for the timely remedy of such issues.

Explainability

Ensure that algorithmic decisions as well as any data driving those decisions can be explained to end-users and other stakeholders in non-technical terms.

Accuracy

Identify, log, and articulate sources of error and uncertainty throughout the algorithm and its data sources so that expected and worst case implications can be understood and inform mitigation procedures.

Auditability

Enable interested third parties to probe, understand, and review the behavior of the algorithm through disclosure of information that enables monitoring, checking, or criticism, including through provision of detailed documentation, technically suitable APIs, and permissive terms of use.

Fairness

Ensure that algorithmic decisions do not create discriminatory or unjust impacts when comparing across different demographics (e.g. race, sex, etc).

We have left some of the terms above purposefully under-specified to allow these principles to be broadly applicable. Applying these principles well should include understanding them within a specific context. We also suggest that these issues be revisited and discussed throughout

the design, implementation, and release phases of development. Two important principles for consideration were purposefully left off of this list as they are well-covered elsewhere: privacy and the impact of human experimentation. We encourage you to incorporate those issues into your overall assessment of algorithmic accountability as well.

Social Impact Statement for Algorithms

In order to ensure their adherence to these principles and to publicly commit to associated best practices, we propose that algorithm creators develop a Social Impact Statement using the above principles as a guiding structure. This statement should be revisited and reassessed (at least) three times during the design and development process:

- design stage,
- pre-launch,
- and post-launch.

When the system is launched, the statement should be made public as a form of transparency so that the public has expectations for social impact of the system.

The Social Impact Statement should minimally answer the questions below. Included below are concrete steps that can be taken, and documented as part of the statement, to address these questions. These questions and steps make up an outline of such a social impact statement.

Responsibility

Guiding Questions

- Who is responsible if users are harmed by this product?
- What will the reporting process and process for recourse be?
- Who will have the power to decide on necessary changes to the algorithmic system during design stage, pre-launch, and post-launch?

Initial Steps to Take

- Determine and designate a person who will be responsible for the social impact of the algorithm.
- Make contact information available so that if there are issues it's clear to users how to proceed
- Develop a plan for what to do if the project has unintended consequences. This may be part of a maintenance plan and should involve post-launch monitoring plans.
- Develop a sunset plan for the system to manage algorithm or data risks after the product is no longer in active development.

Explainability

Guiding Questions

- Who are your end-users and stakeholders?
- How much of your system / algorithm can you explain to your users and stakeholders?
- How much of the data sources can you disclose?

Initial Steps to Take

- Have a plan for how decisions will be explained to users and subjects of those decisions. In some cases it may be appropriate to develop an automated explanation for each decision.
- Allow data subjects visibility into the data you store about them and access to a process in order to change it.
- If you are using a machine-learning model:
 - consider whether a directly interpretable or explainable model can be used.
 - describe the training data including how, when, and why it was collected and sampled.
 - describe how and when test data about an individual that is used to make a decision is collected or inferred.
- Disclose the sources of any data used and as much as possible about the specific attributes of the data. Explain how the data was cleaned or otherwise transformed.

Accuracy

Guiding Questions

- What sources of error do you have and how will you mitigate their effect?
- How confident are the decisions output by your algorithmic system?
- What are realistic worst case scenarios in terms of how errors might impact society, individuals, and stakeholders?
- Have you evaluated the provenance and veracity of data and considered alternative data sources?

Initial Steps to Take

- Assess the potential for errors in your system and the resulting potential for harm to users.
- Undertake a sensitivity analysis to assess how uncertainty in the output of the algorithm relates to uncertainty in the inputs.
- Develop a process by which people can correct errors in input data, training data, or in output decisions.
- Perform a validity check by randomly sampling a portion of your data (e.g., input and/or training data) and manually checking its correctness. This check should be performed early in your development process before derived information is used. Report the overall data error rate on this random sample publicly.
- Determine how to communicate the uncertainty / margin of error for each decision.

Auditability

Guiding Questions

- Can you provide for public auditing (i.e. probing, understanding, reviewing of system behavior) or is there sensitive information that would necessitate auditing by a designated 3rd party?
- How will you facilitate public or third-party auditing without opening the system to unwarranted manipulation?

Initial Steps to Take

- Document and make available an API that allows third parties to query the algorithmic system and assess its response.
- Make sure that if data is needed to properly audit your algorithm, such as in the case of a machine-learning algorithm, that sample (e.g., training) data is made available.
- Make sure your terms of service allow the research community to perform automated public audits.
- Have a plan for communication with outside parties that may be interested in auditing your algorithm, such as the research and development community.

Fairness

Guiding Questions

- Are there particular groups which may be advantaged or disadvantaged, in the context in which you are deploying, by the algorithm / system you are building?
- What is the potential damaging effect of uncertainty / errors to different groups?

Initial Steps to Take

- Talk to people who are familiar with the subtle social context in which you are deploying. For example, you should consider whether the following aspects of people's identities will have impacts on their equitable access to and results from your system:
 - Race
 - Sex
 - Gender identity
 - Ability status
 - Socio-economic status
 - Education level
 - Religion

- Country of origin
- If you are building an automated decision-making tool, you should deploy a fairness-aware data mining algorithm. (See, e.g., the resources gathered at <http://fatml.org>).
- Calculate the error rates and types (e.g., false positives vs. false negatives) for different sub-populations and assess the potential differential impacts.

Authors

Nicholas Diakopoulos, University of Maryland, College Park
 Sorelle Friedler, Haverford College
 Marcelo Arenas, Pontificia Universidad Catolica de Chile, CL
 Solon Barocas, Microsoft Research
 Michael Hay, Colgate University
 Bill Howe, University of Washington
 H. V. Jagadish, University of Michigan
 Kris Unsworth, Drexel University
 Arnaud Sahuguet, Cornell Tech
 Suresh Venkatasubramanian, University of Utah
 Christo Wilson, Northeastern University
 Cong Yu, Google
 Bendert Zevenbergen, University of Oxford

Exhibit 6 - The full text of the Association for Computing Machinery’s “Principles for Algorithmic Transparency and Accountability.”

- 1. Awareness:** Owners, designers, builders, users, and other stakeholders of analytic systems should be aware of the possible biases involved in their design, implementation, and use and the potential harm that biases can cause to individuals and society.
- 2. Access and redress:** Regulators should encourage the adoption of mechanisms that enable questioning and redress for individuals and groups that are adversely affected by algorithmically informed decisions.
- 3. Accountability:** Institutions should be held responsible for decisions made by the algorithms that they use, even if it is not feasible to explain in detail how the algorithms produce their results.
- 4. Explanation:** Systems and institutions that use algorithmic decision-making are encouraged to produce explanations regarding both the procedures followed by the algorithm and the specific decisions that are made. This is particularly important in public policy contexts.
- 5. Data Provenance:** A description of the way in which the training data was collected should be maintained by the builders of the algorithms, accompanied by an exploration of the potential biases induced by the human or algorithmic data-gathering process. Public scrutiny of the data provides maximum opportunity for corrections. However, concerns over privacy, protecting trade secrets, or revelation of analytics that might allow malicious actors to game the system can justify restricting access to qualified and authorized individuals.
- 6. Auditability:** Models, algorithms, data, and decisions should be recorded so that they can be audited in cases where harm is suspected.
- 7. Validation and Testing:** Institutions should use rigorous methods to validate their models and document those methods and results. In particular, they should routinely perform tests to assess and determine whether the model generates discriminatory harm. Institutions are encouraged to make the results of such tests public.

Exhibit 7 - An excerpt from AI Now's April 2018 report "Algorithmic Impact Assessments: A Practical Framework for Public Agency Accountability"

Key Elements of a Public Agency Algorithmic Impact Assessment:

1. Agencies should conduct a self-assessment of existing and proposed automated decision systems, evaluating potential impacts on fairness, justice, bias, or other concerns across affected communities;
2. Agencies should develop meaningful external researcher review processes to discover, measure, or track impacts over time;
3. Agencies should provide notice to the public disclosing their definition of "automated decision system," existing and proposed systems, and any related self-assessments and researcher review processes before the system has been acquired;
4. Agencies should solicit public comments to clarify concerns and answer outstanding questions; and
5. Governments should provide enhanced due process mechanisms for affected individuals or communities to challenge inadequate assessments or unfair, biased, or otherwise harmful system uses that agencies have failed to mitigate or correct.

Exhibit 8 - An excerpt from the overview of Stack Overflow's 2018 Developer Survey Results



Ethics

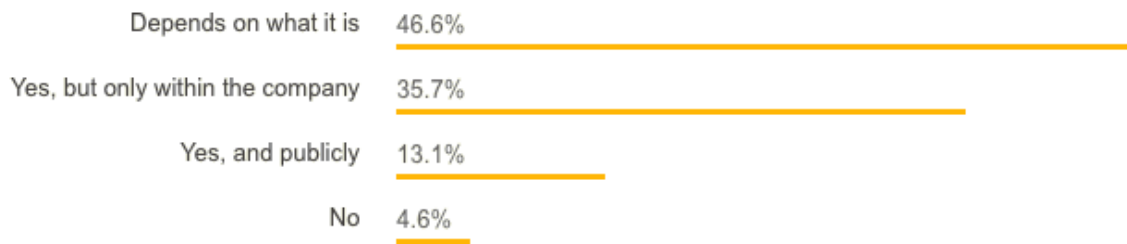
What Would Developers Do If Asked to Write Code for an Unethical Purpose?



70,782 responses

When posed with a hypothetical situation where they are asked to write code for a product or purpose that they consider clearly unethical, over half of our respondents say that they would not write such code. Ethical situations can be complicated, and about another third say that it would depend on the situation.

How Would Developers Report Ethical Problems with Code?



70,426 responses

The question of what to do *next* after an ethical problem has arisen is even more tricky, according to our respondents. The most common answer, from almost half of respondents, is that how to report an ethical problem depends on the particulars of any given situation.

Who Is Ultimately Most Responsible for Code That Accomplishes Something Unethical?



64,540 responses

Most developers feel that management is ultimately most accountable for unethical results of code. Just under 20% of respondents said that a developer who writes code used for unethical purposes is most responsible.

Do Developers Have an Obligation to Consider the Ethical Implications of Their Code?



69,309 responses

Almost 80% of respondents affirm that considering what their code can be used for is the right thing for developers to do. Those who said they were unsure about this were 40% more likely to also say that they do *not* need to report any ethical problems.

We included a free response opportunity after this question, and we saw thoughtful reflections from developers. These include responses about how the tools developers build are powerful and come with a lot of responsibility, situations where unethical outcomes may accidentally arise, and how large teams are involved in building software but developers can be the last line of defense against unethical code.

Exhibit 9 - The full text of New York City’s “algorithmic accountability” bill, as finalized in December 2017

Int. No. 1696-A

By Council Members Vacca, Rosenthal, Johnson, Salamanca, Gentile, Cornegy, Williams, Kallos and Menchaca

A Local Law in relation to automated decision systems used by agencies

Be it enacted by the Council as follows:

Section 1. a. For purposes of this local law:

Agency. The term “agency” means an agency, as defined in section 1-112 of the administrative code of the city of New York, the head of which is appointed by the mayor.

Automated decision system. The term “automated decision system” means computerized implementations of algorithms, including those derived from machine learning or other data processing or artificial intelligence techniques, which are used to make or assist in making decisions.

Automated decision system, agency. The term “agency automated decision system” means an automated decision system used by an agency to make or assist in making decisions concerning rules, policies or actions implemented that impact the public.

Charitable corporation. The term “charitable corporation” shall have the meaning ascribed to such term by section 102 of the not-for-profit corporation law.

b. 1. No later than 120 days after the effective date of this local law, the mayor or a designee thereof shall convene an automated decision systems task force.

2. Such task force and the chair thereof shall be appointed by the mayor or a designee thereof and shall include, but need not be limited to, persons with expertise in the areas of fairness, accountability and transparency relating to automated decision systems and persons affiliated with charitable corporations that represent persons in the city affected by agency automated decision systems, provided that nothing herein shall prohibit the mayor, the designee thereof or the chair from limiting participation in or attendance at meetings of such task force that may involve consideration of information that, if disclosed, would violate local, state or federal law, interfere with a law enforcement investigation or operations, compromise public health or safety or result in the disclosure of proprietary information.

3. No later than 18 months after such task force is established, it shall electronically submit to the mayor and the speaker of the council a report that shall include, at a minimum, recommendations on:

(a) Criteria for identifying which agency automated decision systems should be subject to one or more of the procedures recommended by such task force pursuant to this paragraph;

(b) Development and implementation of a procedure through which a person affected by a decision concerning a rule, policy or action implemented by the city, where such decision was made by or with the assistance of an agency automated decision system, may request and receive an explanation of such decision and the basis therefor;

(c) Development and implementation of a procedure that may be used by the city to determine whether an agency automated decision system disproportionately impacts persons based upon age, race, creed, color, religion, national origin, gender, disability, marital status, partnership status, caregiver status, sexual orientation, alienage or citizenship status;

(d) Development and implementation of a procedure for addressing instances in which a person is harmed by an agency automated decision system if any such system is found to disproportionately impact persons based upon a category described in subparagraph (c);

(e) Development and implementation of a process for making information publicly available that, for each agency automated decision system, will allow the public to meaningfully assess how such system functions and is used by the city, including making technical information about such system publicly available where appropriate; and

(f) The feasibility of the development and implementation of a procedure for archiving agency automated decision systems, data used to determine predictive relationships among data for such systems and input data for such systems, provided that this need not include agency automated decision systems that ceased being used by the city before the effective date of this local law.

4. Such task force shall dissolve 60 days after submission of the report required by paragraph 3.

5. The mayor shall, no later than 10 days after receipt of the report required by paragraph 3, make such report publicly available online through the city's website.

6. Nothing herein shall require compliance with the task force's recommendations or disclosure of any information where such disclosure would violate local, state, or federal law, interfere with a law enforcement investigation or operations, compromise public health or safety, or that would result in the disclosure of proprietary information.

§ 2. This local law takes effect immediately.

MAJ

LS# 10948

12/01/17 2:02 PM

Endnotes

¹ The information in this section comes from both testimony by James Vacca to the New York City Council's Committee on Technology on October 16, 2017, and Lauren Kirchner's interview with him on December 14, 2017.

² Julia Powles, "New York City's Bold, Flawed Attempt to Mark Algorithms Accountable," *The New Yorker*, December 20, 2017, <https://www.newyorker.com/tech/elements/new-york-citys-bold-flawed-attempt-to-make-algorithms-accountable>

³ For more details about how this bill developed over time, including related testimony transcripts and briefing papers, see the New York City Council's web page for it: <http://legistar.council.nyc.gov/LegislationDetail.aspx?ID=3137815&GUID=437A6A6D-62E1-47E2-9C42-461253F9C6D0>

⁴ Testimony by James Vacca to the New York City Council's Committee on Technology, October 16, 2017.

⁵ Jeffrey Baker, "Briefing Paper of the Infrastructure Division," New York City Council Committee on Technology, October 16, 2017.

⁶ Julia Angwin, Jeff Larson, Surya Mattu and Lauren Kirchner, "Machine Bias," *ProPublica*, May 23, 2016, <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing>

⁷ (Technically, it predicted whether a person would be *arrested* for a new crime.)

⁸ Northpointe, Inc. has since changed its name to Equivant.

⁹ The COMPAS questionnaire is available on Document Cloud: <https://www.documentcloud.org/documents/2702103-Sample-Risk-Assessment-COMPAS-CORE.html>

¹⁰ Northpointe disputed ProPublica's analysis, and responded here: <http://www.equivant.com/blog/response-to-propublica-demonstrating-accuracy-equity-and-predictive-parity>

¹¹ Julia Angwin, Jeff Larson, Surya Mattu and Lauren Kirchner, "Machine Bias," *ProPublica*, May 23, 2016, <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing>

¹² Jeffrey Baker, "Briefing Paper of the Infrastructure Division," Committee on Technology, October 16, 2017.

¹³ Solon Barocas and Andrew Selbst, "Big Data's Disparate Impact," *104 California Law Review* 671 (2016). Available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2477899

¹⁴ Executive Office of the President, "Big Data: A Report on Algorithmic Systems, Opportunity, and Civil Rights," *WhiteHouse.gov*, May 2016, https://obamawhitehouse.archives.gov/sites/default/files/microsites/ostp/2016_0504_data_discrimination.pdf

¹⁵ Ibid

¹⁶ Ibid

¹⁷ Another concise list of ways that algorithms' outcomes can seem inaccessible is in the Association for Computing Machinery's 2017 "Statement on on Algorithmic Transparency and Accountability": "Decisions made by predictive algorithms can be opaque because of many factors, including technical (the algorithm may not lend itself to easy explanation), economic (the cost of providing transparency may be excessive, including the compromise of trade secrets), and social (revealing input may violate privacy expectations). Available here: https://www.acm.org/binaries/content/assets/public-policy/2017_usacm_statement_algorithms.pdf

¹⁸ Executive Office of the President, "Big Data: A Report on Algorithmic Systems, Opportunity, and Civil Rights," *WhiteHouse.gov*, May 2016, https://obamawhitehouse.archives.gov/sites/default/files/microsites/ostp/2016_0504_data_discrimination.pdf

¹⁹ Julia Angwin, Jeff Larson, Surya Mattu and Lauren Kirchner, "Machine Bias," *ProPublica*, May 23, 2016, <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing>

²⁰ Ibid

²¹ Ibid

²² Michelle Liu, "Supreme Court refuses to hear Wisconsin predictive crime assessment case," *Milwaukee Journal Sentinel*, June 26, 2017, <https://www.jsonline.com/story/news/crime/2017/06/26/supreme-court-refuses-hear-wisconsin-predictive-crime-assessment-case/428240001/>

- ²³ The court's decision is available on Document Cloud: <https://www.documentcloud.org/documents/2993525-Wisconsin-v-Loomis-Opinion.html>
- ²⁴ Lauren Kirchner, "Wisconsin Court: Warning Labels Are Needed for Scores Rating Defendants' Risk of Future Crime," *ProPublica*, July 14, 2016, <https://www.propublica.org/article/wisconsin-court-warning-labels-needed-scores-rating-risk-future-crime>
- ²⁵ See the court's full decision on Document Cloud: <https://www.documentcloud.org/documents/2993525-Wisconsin-v-Loomis-Opinion.html>
- ²⁶ The COMPAS questionnaire is available on Document Cloud: <https://www.documentcloud.org/documents/2702103-Sample-Risk-Assessment-COMPAS-CORE.html>
- ²⁷ Julia Angwin, Jeff Larson, Surya Mattu and Lauren Kirchner, "Machine Bias," *ProPublica*, May 23, 2016, <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing>
- ²⁸ Julia Angwin and Jeff Larson, "Bias in Criminal Risk Scores is Mathematically Inevitable, Researchers Say," *ProPublica*, December 30, 2016, <https://www.propublica.org/article/bias-in-criminal-risk-scores-is-mathematically-inevitable-researchers-say>
- ²⁹ Ibid
- ³⁰ Sam Corbett-Davies, Emma Pierson, Avi Feller and Sharad Goel, "A computer program used for bail and sentencing decisions was labeled biased against blacks. It's actually not that clear," *The Washington Post*, October 17, 2016, https://www.washingtonpost.com/news/monkey-cage/wp/2016/10/17/can-an-algorithm-be-racist-our-analysis-is-more-cautious-than-propublicas/?utm_term=.450e123cad7e
- ³¹ Alexandra Chouldechova, "Fair prediction with disparate impact: A study of bias in recidivism prediction instruments," presented at the 2016 conference of Fairness, Accountability, and Transparency in Machine Learning (FAT/ML), November 18, 2016. Available here: <https://arxiv.org/abs/1610.07524>
- ³² Sam Corbett-Davies, Emma Pierson, Avi Feller, Sharad Goel, Aziz Huq, "Algorithmic decision making and the cost of fairness," presented at the 23rd SIGKDD Conference on Knowledge Discovery and Data Mining (KDD), August 13-17, 2017. Available here: <https://arxiv.org/abs/1701.08230>
- ³³ Arvind Narayanan, "Translation Tutorial: 21 fairness definitions and their politics." Available here: <https://www.youtube.com/watch?v=jIXIuYdnyyk>
- ³⁴ Ibid
- ³⁵ Narayanan's Twitter thread available here: https://twitter.com/random_walker/status/927625099062243328
- ³⁶ "Shai Danziger, Jonathan Levav, and Liora Avnaim-Pesso, "Extraneous factors in judicial decisions," *Proceedings of the National Academy of Sciences of the United States of America*, Vol. 108, No. 17 (April 26, 2011) pp.6889-6892. Available here: <http://www.pnas.org/content/108/17/6889>
- ³⁷ David Abrams, Marianne Bertrand, and Sendhil Mullainathan, "Do Judges Vary in Their Treatment of Race?" *Journal of Legal Studies*, Vol. 41, No. 2 (June 2012) pp. 347-383. Available here: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1800840
- ³⁸ Ruth Horowitz and Anne E. Pottieger, "Gender Bias in Juvenile Justice Handling of Seriously Crime-Involved Youths," *Journal of Research in Crime and Delinquency*, Vol. 28, Issue 1 (February 1991) pp. 75-100. Available here: <http://journals.sagepub.com/doi/abs/10.1177/0022427891028001005>
- ³⁹ Darrell Steffensmeier, Jeffrey Ulmer and John Kramer, "The Interaction of Race, Gender, and Age in Criminal Sentencing: The Punishment Cost of Being Young, Black, and Male," *Criminology*, Vol. 36, No. 4 (November 1998) pp. 763-798. Available here: <https://onlinelibrary.wiley.com/doi/full/10.1111/j.1745-9125.1998.tb01265.x>
- ⁴⁰ April Michelle Miller, "The Effects of Gender, Race, and Age on Judicial Sentencing Decisions," a thesis presented to the Department of Criminal Justice and Criminology at East Tennessee State University, August 2015. Available here: <https://search.proquest.com/openview/5eacb37e5dede56528e069e8eb9e75da/1?pq-origsite=gscholar&cbl=18750&diss=y>
- ⁴¹ Lauren Kirchner, "For Juvenile Records, It's 'Justice by Geography,'" *Pacific Standard*, November 20, 2014, <https://psmag.com/news/juvenile-records-justice-geography-crime-police-law-enforcement-94909>

- ⁴² Rose Eveleth, “Does Crime Predicting Software Bias Judges? Unfortunately There’s No Data,” *Motherboard*, July 18, 2016, https://motherboard.vice.com/en_us/article/wnxzdb/does-crime-predicting-software-bias-judges-unfortunately-theres-no-data
- ⁴³ Ibid
- ⁴⁴ Ibid
- ⁴⁵ All quotes from this section are taken from the testimony transcript from the October 16, 2017 hearing. Available here: <http://legistar.council.nyc.gov/LegislationDetail.aspx?ID=3137815&GUID=437A6A6D-62E1-47E2-9C42-461253F9C6D0>
- ⁴⁶ Lauren Kirchner, “Thousands of Criminal Cases in New York Relied on Disputed DNA Testing Techniques,” *ProPublica*, September 4, 2017, <https://www.propublica.org/article/thousands-of-criminal-cases-in-new-york-relied-on-disputed-dna-testing-techniques>
- ⁴⁷ For more on the plan to close Riker’s Island, see for instance Reuven Blau and Erin Durkin, “First Rikers Island jail to close in summer as part of city’s 10-year plan to shut down the complex,” *The New York Daily News*, January 2, 2018, <http://www.nydailynews.com/new-york/rikers-jail-close-summer-part-10-year-plan-article-1.3733242> and Lisa W. Foderaro, “New York State May Move to Close Rikers Ahead of City’s 10-Year Timeline,” *The New York Times*, February 14, 2018, <https://www.nytimes.com/2018/02/14/nyregion/rikers-island-jail-closing-timeline.html>.
- ⁴⁸ For more on the Equifax breach, see for instance Thomas Fox-Brewster, “A Brief History Of Equifax Security Fails,” *Forbes*, September 8, 2017, <https://www.forbes.com/sites/thomasbrewster/2017/09/08/equifax-data-breach-history/#6a7e2ff5677c> and Seena Gressin, “The Equifax Data Breach: What to Do,” The Federal Trade Commission, September 8, 2017, <https://www.consumer.ftc.gov/blog/2017/09/equifax-data-breach-what-do>.
- ⁴⁹ Lauren Kirchner interview with James Vacca, October 16, 2017.
- ⁵⁰ Some jurisdictions in the US had previously passed laws specifically governing the testing and the use of algorithms in the criminal justice field, but no other jurisdiction had passed a law providing a framework for *all* algorithms in government.
- ⁵¹ More details about the GDPR can be found on the European Commission website: https://ec.europa.eu/commission/priorities/justice-and-fundamental-rights/data-protection/2018-reform-eu-data-protection-rules_en
- ⁵² The 2018 AI Now report “Algorithmic Impact Assessments” has more on how the GDPR can provide a useful starting point for algorithmic-accountability policy discussions, with caveats: “The GDPR language may be a good starting point for some agencies, but will require some shaping to match the appropriate contexts.” See Dillon Reisman, Jason Schultz, Kate Crawford, Meredith Whittaker, “Algorithmic Impact Assessments: A Practical Framework for Public Agency Accountability,” April 2018, <https://ainowinstitute.org/aiareport2018.pdf>
- ⁵³ Danielle Keats Citron, “Big Data Should Be Regulated by “Technological Due Process,”” *The New York Times*, July 29, 2016, <https://www.nytimes.com/roomfordebate/2014/08/06/is-big-data-spreading-inequality/big-data-should-be-regulated-by-technological-due-process>
- ⁵⁴ Lauren Kirchner, “When Big Data Becomes Bad Data,” *ProPublica*, September 2, 2015, <https://www.propublica.org/article/when-big-data-becomes-bad-data>
- ⁵⁵ *Griggs v. Duke Power Company*, decided March 8, 1971. Court opinion available here: <https://caselaw.findlaw.com/us-supreme-court/401/424.html>
- ⁵⁶ *Texas Department of Housing and Community Affairs v. The Inclusive Communities Project, Inc.*, decided June 25, 2015. Court opinion available here: <https://supreme.justia.com/cases/federal/us/576/13-1371/>
- ⁵⁷ David Ingold and Spencer Soper, “Amazon Doesn’t Consider the Race of Its Customers. Should It?” *Bloomberg*, April 21, 2016, <https://www.bloomberg.com/graphics/2016-amazon-same-day/>
- ⁵⁸ Julia Angwin and Jeff Larson, “The Tiger Mom Tax: Asians Are Nearly Twice as Likely to Get a Higher Price from Princeton Review,” *ProPublica*, September 1, 2015, <https://www.propublica.org/article/asians-nearly-twice-as-likely-to-get-higher-price-from-princeton-review>
- ⁵⁹ Solon Barocas and Andrew Selbst, “Big Data’s Disparate Impact,” *104 California Law Review* 671 (2016). Available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2477899

⁶⁰ The ACLU’s brief is available at <https://www.aclu.org/legal-document/california-v-johnson-amicus-brief> and the EFF’s is at https://www.eff.org/files/2017/09/14/f071640_filed_copy_johnson_amicus_brief_9.13.17.pdf. See also Lauren Kirchner, “Where Traditional DNA Testing Fails, Algorithms Take Over,” *ProPublica*, November 4, 2016, <https://www.propublica.org/article/where-traditional-dna-testing-fails-algorithms-take-over>

⁶¹ Lauren Kirchner, “Federal Judge Unseals New York Crime Lab’s Software for Analyzing DNA Evidence,” *ProPublica*, October 20, 2017, <https://www.propublica.org/article/federal-judge-unseals-new-york-crime-labs-software-for-analyzing-dna-evidence>

⁶² Lauren Kirchner, “Thousands of Criminal Cases in New York Relied on Disputed DNA Testing Techniques,” *ProPublica*, September 4, 2017, <https://www.propublica.org/article/thousands-of-criminal-cases-in-new-york-relied-on-disputed-dna-testing-techniques>

⁶³ See ProPublica’s GitHub page: <https://github.com/propublica/nyc-dna-software>

⁶⁴ EPIC’s project is available at <https://epic.org/algorithmic-transparency/crim-justice/> and MuckRock’s is at <https://www.muckrock.com/project/uncovering-algorithms-84/>.

⁶⁵ Ellora Israni, “Algorithmic Due Process: Mistaken Accountability and Attribution in *State v. Loomis*,” *JOLT Digest*, August 31, 2017, <https://jolt.law.harvard.edu/digest/algorithmic-due-process-mistaken-accountability-and-attribution-in-state-v-loomis-1>

⁶⁶ Dan Hurley, “Can an Algorithm Tell When Kids Are in Danger?” *The New York Times Magazine*, January 2, 2018, <https://www.nytimes.com/2018/01/02/magazine/can-an-algorithm-tell-when-kids-are-in-danger.html>

⁶⁷ *Ibid*

⁶⁸ Executive Office of the President, “Big Data: A Report on Algorithmic Systems, Opportunity, and Civil Rights,” *WhiteHouse.gov*, May 2016, https://obamawhitehouse.archives.gov/sites/default/files/microsites/ostp/2016_0504_data_discrimination.pdf

⁶⁹ “Principles for Accountable Algorithms and a Social Impact Statement for Algorithms,” available here: <https://www.fatml.org/resources/principles-for-accountable-algorithms>

⁷⁰ *Ibid*

⁷¹ Association for Computing Machinery US Public Policy Council (USACM), “Statement on Algorithmic Transparency and Accountability,” January 12, 2017, https://www.acm.org/binaries/content/assets/public-policy/2017_usacm_statement_algorithms.pdf

⁷² Alex Campolo, Madelyn Sanfilippo, Meredith Whittaker, and Kate Crawford, “AI Now 2017 Report,” January 2017, https://ainowinstitute.org/AI_Now_2017_Report.pdf

⁷³ Dillon Reisman, Jason Schultz, Kate Crawford, Meredith Whittaker, “Algorithmic Impact Assessments: A Practical Framework for Public Agency Accountability,” April 2018, <https://ainowinstitute.org/aiareport2018.pdf>

⁷⁴ Bettina Berendt, Kate Crawford, Jon Kleinberg, Hanna Wallach, and Suresh Venkatasubramanian (moderator), “Closing panel: Building a community and setting a research agenda,” 2016 FAT/ML Conference, November 18, 2016. Available here: <https://www.fatml.org/schedule/2016/presentation/closing-panel-building-community-and-setting-resea>

⁷⁵ *Ibid*

⁷⁶ The 2018 StackOverflow survey results are available here: <https://insights.stackoverflow.com/survey/2018/>

⁷⁷ Lauren Kirchner, “New York City Moves to Create Accountability for Algorithms,” *ProPublica*, December 18, 2017, <https://www.propublica.org/article/new-york-city-moves-to-create-accountability-for-algorithms>

⁷⁸ Julia Powles, “New York City’s Bold, Flawed Attempt to Mark Algorithms Accountable,” *The New Yorker*, December 20, 2017, <https://www.newyorker.com/tech/elements/new-york-citys-bold-flawed-attempt-to-make-algorithms-accountable>

⁷⁹ *Ibid*

⁸⁰ All information and quotes from Rashida Richardson in this section come from Lauren Kirchner’s interview with her, December 15, 2017.

⁸¹ All information and quotes from James Vacca in this section come from Lauren Kirchner’s interview with him, December 14, 2017.



This work is licensed under a [Creative Commons Attribution 4.0 International License](https://creativecommons.org/licenses/by/4.0/).