Ghosts in the Machine: Algorithmic Bias and the Courts

By John G. Browning and Alex Shahrestani

I. INTRODUCTION

When asked if he could foresee a day “when smart machines, driven with artificial intelligence, will assist with courtroom fact-finding or, more controversially even, judicial decision-making,” U.S. Supreme Court Chief Justice John Roberts gave a startling response. He said “It’s a day that’s here and it’s putting a significant strain on how the judiciary goes about doing things.”

And despite how many might reserve the notion of using technology to predict an individual’s likelihood to commit crimes for dystopian science fiction movies like Steven Spielberg’s Minority Report, the simple truth is that it is already here. In many ways, much of our modern lives is already impacted by machine learning—the intelligent algorithms underlying the artificial intelligence revolution that already determines everything from Facebook newsfeeds and Google search results to Netflix viewing and online shopping recommendations. Algorithms can also help determine whether we get approved for a mortgage or a job interview, as well as what penalties we face if we commit a crime.

But while the success and widespread use of intelligent algorithms have led many to adopt a “numbers do not lie” attitude and imbue such machine learning with what researchers Osonde Osoba and William Wesler IV describe as an “aura of objectivity and infallibility,” the truth is that certain algorithms utilizing machine learning tools do mimic human biases. And while the shortcoming of some built-in bias—the lingering “ghost in the machine” unconsciously left by a programmer—may seem trivial when the consequences are limited to online dating matchups, it becomes far weightier when crime and punishment are concerned. Skewed input data, imperfect or false logic, or just the prejudices of their programmers can result in intelligent algorithms responding (and in some cases, amplifying) human biases.

In the justice system, a number of legal challenges to algorithmic bias in various contexts have already taken place. In the employment arena, the Houston public school teachers union challenged the use of proprietary algorithms for school employment practices. Facebook

---


settled an Equal Employment Opportunity Commission (EEOC) complaint alleging that Facebook enabled gender and age-based discrimination in hiring practices through the use of the social networking platform’s ad-targeting features, which allowed for job postings to be displayed to audiences narrowed by age and gender. Minorities disproportionately impacted by credit and mortgage lending decisions have brought lawsuits alleging algorithmic bias. In the case *K.W. v. Armstrong*, the American Civil Liberties Union (ACLU) of Idaho brought a class action lawsuit on behalf of about 4,000 individuals with developmental and intellectual disabilities challenging the state’s use of a “black box” algorithm to make Medicaid benefits determinations.3

This article, however, will focus on algorithmic bias in the criminal justice context. We will begin with an introduction to how such machine learning works generally, as well as a look at the types of algorithmic bias. The article will then examine some examples of algorithmic bias, from the “predictive policing” initiatives of cities like Chicago and Oakland, to challenges to algorithmic bias in cases like 2016’s *State v. Loomis*. Finally, we will discuss some of the proposed measures to combat algorithmic bias, including federal legislation like the Algorithmic Accountability Act. Overall, as we shall see, an intelligent algorithm is only as good as the data it learns from. Machine learning on inherently biased data cannot help but lead to biased results.

**II. HOW MACHINE LEARNING WORKS**

When you were a baby, just learning how to talk, you observed thousands of conversations. You noticed that there was a cadence to the sounds being expressed, you noticed that some of those sounds were repeated over and over again, and you learned your first words. After thousands of hours of training, you pumped out a single word. From there, you probably started trying to string together “sentences” in a series of meaningless babbling, with an occasional coherent word thrown in. The sentence did not achieve the desired outcome, and you tried again. Over time and much trial and error, you learned to talk.

Machine learning works in much the same way. A program attempts to achieve an outcome by modeling its outputs against the data you provide it. To carry the analogy forward, you provide the program with thousands of hours of speech recordings for it to listen and model itself

---

after. The program tries to match cadence with meaning, and it eventually “learns” to synthesize speech.

However, imagine that instead of providing the program with thousands of hours of actual speech recordings, you provided the program with mostly speech recordings, but maybe a quarter of the recordings were of chattering monkeys. As far as the computer is concerned, all of the recordings are equally valid, so it will attempt to model its synthesized speech after the entirety of the recordings, resulting in sentences that include the occasional monkey-based vocalization or “speech” structure. This would be an obvious error to any person listening to the synthesized speech, but the use of machine learning is often to remove the need for an individual to interact with the data: the program is likely to go on and on synthesizing monkey sounds into the outputs long before anyone notices.

This can lead to many problems when the machine learning tool is responsible for important decisions: (1) whom should receive a job interview; (2) whom should receive loan money; (3) what stocks to invest in; or (4) whom should be sent to jail. All of those listed use cases to rely on the data fed into it, and the data can be problematic for a few reasons.

III. TYPES OF MACHINE LEARNING BIAS

There are various kinds of machine learning bias; or rather, there are various ways in which machine learning bias manifests itself. As a result there are three avenues by which machine learning bias can occur, and therefore must be addressed.

A. Pre-Existing

Pre-existing machine learning bias is the codification of already-present biases. If a system designer has a real prejudice that he/she wants to implement into a technological solution, that would be an instance of pre-existing machine learning bias. Another such instance would be the inclusion of an implicit bias into the system, something that the system designer is not cognizant of as a cause of discrimination. Pre-existing machine learning bias means a bias that would exist regardless of the machine learning solution, but the machine learning incorporates that bias into its processes.

B. Technical

Technical machine learning bias is the bias that occurs due to the technical limitations of actually presenting the data. If an employer is presented top candidates for a position in a structured order not based on scoring, then candidates are either going to be advantaged or disadvantaged – the first name on a list of top candidates will have a significant advantage...
over those at the bottom of the list. Another example could be that the data gathering mechanism is most robust on the most advanced phones on the market, and the others have a reduced data-set. Technical bias problems reflect a serious difficulty in being objective in presenting results.

C. Emergent
Emergent machine learning bias is the development of new biases or new understandings of biases as technology develops. For example, if audiobooks became so popular a method of consuming literature that published books were made obsolete, then the deaf population would be negatively impacted. A different example would be the development of a new trend in society that has not been accounted for in creating processes for sorting big data—such as a demographic survey not reflecting third options for gender identifiers following a changing social awareness around gender identities.

IV. Algorithmic Bias in the Courts
A. Predicting Risk
Using mathematics to guide decision-making in the criminal justice system is hardly a new phenomenon. As far back as the 1920s, the Illinois parole board used mathematical tabulations that assessed risk by comparing people up for parole to offenders who had already been released. But while the mathematics behind those tools has improved and statisticians can now grapple with far greater data sets using computers, the problem of inherent bias remains — particularly in the use of predictive analytics. For example, in 2016, the independent investigative journalism project ProPublica studied the “risk scores” and assessments of Northpointe, Inc.—the Michigan–based company behind the Correctional Offenders Management Profiling for Alternative Sanctions (COMPAS)—for 7,000 people who were arrested in Broward County, Florida. These scores are used to determine release dates and bail, since they supposedly predict the defendant’s likelihood to commit a crime again. In the cases investigated, ProPublica maintains the algorithms wrongly labeled African American defendants as future criminals at a rate nearly twice (43% vs. 23%) that of white defendants (who were far more likely to be labeled as “low risk” than black defendants). ProPublica’s study also found that the COMPAS algorithm was 61% predictive of re-arrest, or “somewhat more accurate than a coin flip.” In addition, only 20% of the people predicted to commit violent crimes actually went on to do so. Besides just the errors alone, ProPublica found the software

4 Jeff Larsen et al., How We Analyzed the COMPAS Recidivism Algorithm, ProPublica (May 23, 2016), https://www.propublica.org/article/how-we-analyzed-the-compas-recidivism-algorithm.
to be biased against African Americans, with blacks given what would appear to be an inflated risk score despite the input of facts suggesting the opposite. In one case, an African American woman with four juvenile misdemeanors was given a score of 8, while a white male with 2 armed robberies and an attempted armed robbery was assessed a score of just 3. While ProPublica’s thought-provoking finding generated considerable media attention and raised many questions, it could only go so far due to a persistent issue with such algorithms: transparency. Unable to analyze the algorithm itself because its corporate owners refused to release the code, ProPublica could only rely on input and output data.

A study involving the City of Oakland provides another look at algorithmic bias. In 2016, the Human Rights Data Analysis Group investigated PredPol, an algorithm already in use in several states and designed to predict when and where crimes will take place. Its analysis found that rather than accurately predicting crime, the algorithm exacerbated past racially biased policing practices by leading police to target certain neighborhoods. Applying the algorithm to drug offenses in Oakland, California, the study found that officers were repeatedly sent to predominantly black neighborhoods—areas disparately overrepresented in prior arrest data. PredPol responded that this was merely training data, and that in real world applications the software is not used to predict drug crime.

The Chicago Police Department (CPD) has also turned towards a predictive policing initiative in order to reduce gun violence. It is using a technology that takes various, unknown factors into account, runs those factors through an algorithm, and scores people as part of a “heat list.” Oftentimes, those on the heat list are then directly contacted by the CPD to notify them that they are on the CPD’s radar as people to watch. The people on the heat list are either invited to a community meeting, notified through communications, or are told in person at their homes. The software’s variables as well as the maker of the software are completely unknown and unexamined; all that has been revealed is that criminal history, known criminal associates, and whether one has been the victim of a crime are somehow included in the process. Given the serious nature of the consequences of a computer program determining who is most likely to become a criminal, it was inevitable that a lawsuit would emerge in order to determine the underlying processes for the program. The Chicago Sun-Times filed a lawsuit in Cook County’s Court of Chancery under the Freedom of Information Act to find out the nature of the algorithm, the maker of the algorithm, and the race of each person on the list, among other

factors. The CPD refused the initial FOIA request, claiming it would be “unduly burdensome” to provide those details.

B. State of Wisconsin v. Loomis

During the 1990s, the company Northpointe, Inc. worked on the development of COMPAS, an intelligent algorithm designed to assess the risk that a given defendant will commit a crime after release. It uses a number of factors, including a defendant’s own responses to a lengthy questionnaire, to generate a recidivism risk score between 1 and 10 by comparing a given defendant’s traits to those of known high-risk offenders. It then classifies the risk of recidivism as low risk (1–4), medium risk (5–7), or high risk (8–10). The score is then included as part of a defendant’s presentence investigation (PSI) report for the sentencing judge.

In 2012, Wisconsin implemented COMPAS into its state sentencing procedures. In 2013, 35 year–old Eric Loomis was arrested for his involvement in a drive–by shooting in La Crosse, Wisconsin. No one was hurt, but Loomis was driving the getaway vehicle, a stolen car. He pled no contest to two lesser charges—“attemping to flee a traffic officer” and “operating a motor vehicle without the owner’s consent.” The trial judge sentenced Loomis to 7 years, based in part on a COMPAS score assessing him as a “high risk.” Loomis filed a motion for post–conviction relief seeking a new sentencing hearing, arguing that the court’s consideration of the COMPAS risk assessment violated his constitutional rights to due process. He further argued that the trial court erred by improperly assuming that the factual bases for the risk assessment were true.

The case went all the way to the Wisconsin Supreme Court, as Loomis challenged the lack of transparency with the algorithm used to sentence him. Loomis argued that while the sentencing judge could view the risk score itself and the inputs affecting it, no one—not even the judge—knew what decisions the software had been programmed to make. Loomis

7 A New York Times analysis revealed that the likely most significant factors (among known factors) in being a high–risk subject were the number of times one was the victim of a violent crime, with the biggest counter–weight being age. Jeff Asher & Rob Arthur, Inside the Algorithm That Tries to Predict Gun Violence in Chicago, N.Y. TIMES (June 13, 2017), https://www.nytimes.com/2017/06/13/upshot/what–an–algorithm–reveals–about–life–on–chicagos–high–risk–list.html?_r=1.
9 State v. Loomis, 881 N.W.2d 749 (Wis. 2016).
contended that Northpointe (and the software company that had written the algorithm, Equivant) should be required to divulge its source code. Because the companies steadfastly refused to do so, citing its proprietary nature and invoking the trade secrets privilege, Loomis asserted that because the scientific validity of the tool could not be determined, his due process rights had been violated. As an expert for Loomis testified, “There’s all kinds of information that the court doesn’t have,” and because too little is known about how the risks are analyzed, “COMPAS should not be used for incarceration decisions.”

But the Wisconsin Supreme Court disagreed, stating “[W]e conclude that if used properly, observing the limitations and cautions set forth herein, a circuit court’s consideration of a COMPAS risk assessment at sentencing does not violate a defendant’s right to due process.”

The court also considered that because COMPAS uses only publicly available data and data provided by the defendant himself, Loomis could have denied or explained any information that went into the making of the report.

However, the Court also ruled that courts should proceed with caution, and not make an assessment report the sole basis for its sentencing decision. It further held that the use of a COMPAS risk assessment must be subject to certain cautions. Specifically, PSIs accompanying COMPAS assessments must include five written warnings for judges: (1) that the “proprietary nature of COMPAS” prevents the disclosure of how risk scores are calculated; (2) that COMPAS scores are unable to identify specific high-risk individuals because the scores themselves rely on group data; (3) that although COMPAS relies on a national data sample, there has been “no cross-validation study for a Wisconsin population”; (4) that studies have “raised questions” about whether COMPAS scores disproportionately classify minority offenders as having a higher risk of recidivism; and (5) that COMPAS was developed specifically for a different purpose—to assist the Department of Corrections in making post-sentencing determinations.

In a concurring opinion, Justice Shirley Abrahamson expressed concern with “the court’s lack of understanding of COMPAS,” calling it a “significant problem” and bemoaning the fact that “few answers were available” to the questions that judges had directed to Northpointe as the case wound its way through the courts. Greater explanation was needed, Justice Abrahamson wrote,

---

10 Id.
11 Id. at 761–62.
12 Id. at 769–70.
in part because “the use of risk assessment tools like COMPAS has garnered mixed reviews in
the scholarly literature and in popular commentary and analysis.”13

The Loomis opinion, with its urging of caution, its careful distinction that a court may
“consider” rather than “rely” on such risk assessments, and its requirement of written
disclaimers for PSIs suggests that perhaps enthusiasm over algorithmic risk assessments will
be tempered in the future. Only time will tell.

C. Rodriguez and Beyond

The court in Loomis justified its decision in part by noting that, since COMPAS uses only
publicly available data and data provided by the defendant, Loomis could have denied or
explained any information that went into making the report and accordingly could have verified
the accuracy of the information used in sentencing. But in another case involving COMPAS, the
defendant had a much more challenging task in denying or explaining the information that
provided the basis for the report. In 2017, Glenn Rodriguez, an inmate at the Eastern
Correctional Facility in upstate New York with a virtually spotless rehabilitation record, was
denied parole due to a “high risk” COMPAS score. The bewildered Rodriguez consulted with
lawyers and was able to review the 137-item questionnaire that had been filled out about him
and which comprised the basis for his COMPAS score. Rodriguez found a mistake in an answer
given by the correctional officer who had filled out the form, and learned that other inmates
had nearly identical questionnaires to him, but with entirely different scores. Although he had
no access to the methodology of the COMPAS assessment, at his second parole hearing in
January 2017, Rodriguez argued that since the input was wrong, his final risk score could not
possibly be accurate. Yet because he did not know how the information was weighted in the
algorithm’s code, he could not prove how significant the error was. Rodriguez was ultimately
granted parole in May 2017, but without being able to challenge the troubling issues of his
COMPAS risk score.14

COMPAS remains the most widely used algorithm for risk assessment. At least nine states—
Arizona, Colorado, Delaware, Kentucky, Louisiana, Oklahoma, Virginia, Washington, and
Wisconsin—use its assessments during criminal sentencing hearings. And despite continued
concerns about accuracy, the lack of transparency, and of course the perpetuation of biases,
risk assessment algorithms like COMPAS are still used at multiple junctures throughout the

13 Id. at 774–75.
criminal justice process: (1) to determine the length of sentences; (2) to direct defendants to alternatives to incarceration like rehabilitation; and (3) to shorten sentences for good behavior.

If courts are not going to be receptive to challenges of intelligent algorithms because they are apparently more protective of a company’s intellectual property rights in its proprietary software than of an individual defendant’s due process rights, what about challenges to algorithms that are based on product liability theories? In the first case of its kind, a federal court in New Jersey considered a product liability claim brought against the owners of an algorithm risk assessment tool. Plaintiff June Rodgers sued over the death of her 26-year-old son Christian Rodgers at the hands of Jules Black on April 9, 2017. Black had been arrested on April 5, 2017, by New Jersey State Police and charged with being a felon in possession of a firearm. According to the lawsuit, Black was released on non-monetary conditions the next day because he had a low Public Safety Assessment (PSA) score, pursuant to the defendant’s algorithm. Mrs. Rodgers brought suit under the New Jersey Product Liability Act, arguing that the algorithm had turned out a low PSA for Black because it was a defective product.

The court, however, disagreed. It held that the PSA is not a product as defined by New Jersey’s statute. Moreover, the judge reasoned, the algorithm is “neither a tangible product or a non-tangible other item as contemplated by section 19 of the Restatement of Torts and it is not distributed commercially.” Ruling that the PSA “constitutes information, guidance, ideas, and recommendations as to how to consider the risk a given criminal defendant presents,” the court reasoned that the PSA could not be subject to tort liability since it would be properly treated “as speech, rather than product.” The court also rejected Rodgers’ contention that such PSAs “thwart” the role of judges, noting that “the PSA does not supplant judicial decision making but merely informs a judge’s decision of whether to release or detain a defendant pending trial.”

---

15 Risk assessment algorithms are not the only tech tool being challenged by criminal defendants. The makers of TrueAllele, a software program used to analyze traces of DNA from crime scenes, have faced legal action from defendants seeking to review its source code in order to confront and cross-examine its programmer about how the software works. Like Northpointe, TrueAllele’s developers have successfully relied on trade secret evidentiary privilege to quash such attempts at discovery. See, e.g., People v. Chubbs, 2015 WL 139069 (Cal. App. June 9, 2015).
17 Id.
18 Id.
19 Id.
V. MEASURES ADDRESSING ALGORITHMIC BIAS

On the international level, the European Union has already taken steps to address the fallibility of algorithms. Article 22 of the General Data Protection Regulation (GDPR) protects most people from wholly automated decision-making processes, requiring (in many circumstances) human review of appealed automated decisions—even where wholly autonomous decision-making processes are permitted. On a far more local level, the New York City Council reacted to ProPublica’s study by establishing the Automated Decision System Task Force in 2017 to study and make recommendations of policies, practices, and guidelines on the use of such systems in all city-wide public agencies. The task force hopes to curb algorithmic bias and promote transparency, including in the area of criminal sentencing.

During the Obama Administration, an effort at introducing an algorithmic accountability law as part of the Consumer Privacy Bill of Rights failed. In April 2019, Senators Ron Wyden (D-Ore.) and Cory Booker (D-N.J.) introduced their own Algorithmic Accountability Act. Under it, the Federal Trade Commission would compel companies to test both their algorithms and training data for any defects that could lead to biased, inaccurate, discriminating, or otherwise unfair decisions. Companies would be required to assess the objectivity of their algorithms, and to address any flaws uncovered during the assessment. In addition, entities subject to the Act would need to ensure that they are protecting the privacy and security of the consumer data being fed into the algorithms. The Act would apply to companies that make over $50 million a year, hold information on at least one million people or devices, or which primarily act as data brokers buying and selling consumer data. The bill was introduced just weeks after Facebook was sued by the Department of Housing and Urban Development, which alleged that the site’s ad targeting tools unfairly discriminated on the basis of gender and race in determining who would see certain housing advertisements. As Senator Wyden noted in introducing the bill:

computers are increasingly involved in the most important decisions affecting American lives—whether or not someone can buy a home, get a job or even go to jail. But instead of eliminating bias, too often these algorithms depend on biased assumptions or data than can actually reinforce discrimination against women and people of color.

But the proposed Act is geared toward consumer protection, not protection of due process and other rights of criminal defendants. Perhaps the best way to address algorithmic bias is not to

---

be found in legislative solutions at all, but in improving the technology itself. The Laura and John Arnold Foundation, a Houston-based philanthropic organization, has developed the Public Safety Assessment Court tool, an algorithm used in bail decisions. Currently used in thirty jurisdictions, the PSA court tool is not “black boxed” and does not rely on socio-economic factors or gender. Instead, it considers nine factors related to a person’s criminal history before providing a risk assessment on how likely the person is to fail to appear for a court date, or commit a new crime or violent crime while on release. The factors include prior misdemeanor and felony convictions, previous failures to appear, and the defendant’s age. Studies of the tool are encouraging; research into its use in Ohio, for example, found that outcomes did not show any race or gender bias. The number of people released without the need for bail doubled from 14% to nearly 28%. The percentage of pretrial defendants arrested for other crimes while on bail was cut in half (from 20% to 10%), and the percentage of those arrested for violent crimes while out on bail also dropped (from 5% to 3%). The Arnold Foundation touts not only these results but also the transparency of the PSA court tool.

VI. CONCLUSION

In a 2014 address to the National Association of Criminal Defense Lawyers, then-U.S. Attorney General Eric Holder warned of the dangers of algorithmic bias:

> Although these measures were crafted with the best of intentions, I am concerned that they may inadvertently undermine our efforts to ensure individualized and equal justice . . . By basing sentencing decisions on static factors and immutable characteristics—like the defendant’s education level, socioeconomic background, or neighborhood—they may exacerbate unwarranted and unjust disparities that are already far too common in our criminal justice and in our society.

While the use of intelligent algorithms for criminal risk assessment has been promoted as a means of eliminating human bias in sentencing, the fact is that methodologies that are necessarily created by individuals may reflect human bias. When data points that may be race neutral on their face—such as zip codes or family history of incarceration—but which can serve as proxies for race are employed, human bias is not only not eliminated, it is automated. The Loomis case and others discussed in this article illustrate this danger, particularly when courts place a higher priority on the protection of intellectual property rights—the “black box” or proprietary “secret sauce” of an intelligent algorithm. Without greater transparency into the workings of a given algorithm, the opaque nature of letting machines function as judge and
jury means no one knows if poor results are being produced until the damage has already been done.

But there is hope. As the field of “algorithmic fairness” grows, data scientists are developing ways to identify and correct disparate impact in machine learning algorithms. After all, part of the problem lies in the reiterative nature of algorithms, many of which “learn” from themselves by running repeatedly, studying and reapplying the results, and compounding them. When even a small amount of bias is introduced, such an algorithm cannot help but reproduce it in a troubling feedback loop.

The complexity of achieving justice rests in the balancing of many factors—deterrence, fairness, proportionality, empathy, victims’ and society’s calls for punishment, to name just a few. While algorithms can illuminate these goals, they should not be the determinative factor in decisions impacting individual rights.

About the Authors

John Browning is an attorney in Dallas who litigates a wide variety of civil litigation in state and federal courts throughout Texas, including commercial disputes, personal injury and wrongful death defense, employment matters, health care, and intellectual property litigation. He is the author of the books The Lawyer’s Guide to Social Networking, Understanding Social Media’s Impact on the Law, (West 2010); the Social Media and Litigation Practice Guide (West 2014); Legal Ethics and Social Media: A Practitioner’s Handbook (ABA Press 2017); and Cases & Materials on Social Media and the Law (forthcoming). Mr. Browning is also a contributing author to seven other books, the author of nearly 35 published law review articles; and the award–winning writer of numerous articles for regional and national legal publications. His work has been cited in nearly 350 law review articles, practice guides in 11 states, and by courts in Texas, California, Maryland, Tennessee, New York, Florida, Illinois, and Puerto Rico. He has been quoted as a leading authority on social media and the law by such publications as The New York Times, The Wall Street Journal, USA Today, Law 360, Time Magazine, The National Law Journal, the ABA Journal, WIRED Magazine and Inside Counsel Magazine, and he is a recurring legal commentator for the NBC, CBS, and FOX news stations in Dallas. He is an adjunct professor at SMU Dedman School of Law and he serves as the Chair of the Computer & Technology Section of the State Bar.
Alex Shahrestani is a startup-tech nerd trapped in an attorney’s body. He serves as Vice President of EFF–Austin, CLE Program Coordinator for SXSW, a leadership member of the Computer & Technology Section of the State Bar, a leadership member of Texas Exes Young Alumni– Austin, and the Founder of the Journal of Law and Technology at Texas. His practice focuses on startup and small business issues, and he provides subscription services for his clients. You can find out more about him and how he uses his CS background to inform his practice at shahrestanilaw.com.