CODING OUT IMPLICIT BIAS WITH AI

By Sarah Kellogg

No longer just a powerful productivity tool in law firms, artificial intelligence (AI) and smart algorithms are being enlisted to combat one of society's most entrenched and thorny problems: the implicit bias found in human decision making.

The promise of the latest applications of AI is that it can somehow wash away the hidden bias that too frequently pervades institutional judgments, such as weighing variables to determine whether certain individuals are eligible for a job, a bank loan, or parole.

The potential here is enormous, the promise yet unproven, say observers.

Earlier this year, San Francisco District Attorney George Gascón announced that his office would use AI to scan police reports and remove all references to race, allowing prosecutors to make race-free decisions when initially charging individuals of crimes.

"Lady Justice is depicted wearing a blindfold to signify impartiality of the law, but it is blindingly clear that the criminal justice system remains biased when it comes to race," said Gascón in a written statement announcing the AI deployment. "This technology will reduce the threat that implicit bias poses to the purity of decisions."

But AI's critics have urged caution, noting that the same biases that creep into human decision making can find their way into programs and algorithms, and the consequences could be far more widespread and long lasting.

"When you start to automate with AI, you are outsourcing decision making to the people who created that program, and you cannot know what data influenced their decisions," says Alexandra Givens, an adjunct professor of law and executive director of Georgetown University Law Center's Institute for Technology Law & Policy. "It is one thing to have a single biased judge, but the decisions coded into a hiring platform will apply to every candidate who goes through that platform, making it a potentially biased system."

The legal industry has already embraced AI-driven, productivity-focused software and applications, harnessing the power of technology to master e-discovery, predictive analytics, and contract review. Up to this point, AI's central value has been its ability to automate and streamline processes that are considered arduous and labor intensive inside law firms.

AI providers are hoping to eliminate the deeply ingrained biases and attitudes that affect an individual's unconscious understanding and actions. While these can be both positive and negative biases, they are generally activated involuntarily and without warning. They cannot be reduced through introspection, which leaves researchers to elevate an algorithmic solution.

Whether the legal industry will fully embrace this latest incarnation of technology depends largely on its promoters and whether they can convince law firm leaders and criminal justice officials that objective algorithms are better at executing decisions than biased humans.

A NOVEL HR SOLUTION FOR FIRMS

The first U.S. law firm to pioneer bias-free algorithms in the recruitment process is O'Melveny & Myers LLP, which announced in November 2018 that it would use neuroscience-infused online games to assess associate applicants for a range of cognitive, social, and emotional traits.

"A number of corporations have used these types of games, but we are the first law firm to use them," says Darin Snyder, O'Melveny's diversity and inclusion partner and member of the firm's executive committee. "The legal profession is sometimes slow to innovate. In this context, we're happy to be that brave soul that goes first."

The law firm hired the New York-based company pymetrics to design games to meet O'Melveny's specific requirements. First, the firm had its high-performing associates play the games to generate the training data necessary to "educate" the algorithm. AI measured the O'Melveny team's levels of effort, attention, planning, memory, and flexibility. Using these results, and removing any gender, race, or ethnic identifiers, pymetrics built a success profile for the games.

When law school students play the games this year in hopes of securing an interview with O'Melveny, the algorithm will be able to cull the best candidates. The focus is on prospective success and not on pedigree, says Snyder.

Companies such as Unilever, Accenture, and LinkedIn also use games in their recruitment process.

"These games are not like any personality test that you've ever taken," says Snyder. "They don't ask subjective questions. They are actual games that are based on tests used in sociological literature for decades. Millions of people have played these types of games, and we know as a result that they are predictive."

One of the deciding factors for O'Melveny was that pymetrics provided evidence the games are free of bias and will not institutionalize any historic bias. Law firm leaders also liked the fact that pymetrics created a student report of the findings, which can be shared with other potential employers that pymetrics serves. To increase impartiality, pymetrics only offers O'Melveny three potential verdicts for each game player: highly recommend, recommend, and not recommend.

Before implementing the new AI tools, O'Melveny hosted a series of calls this past year with the law schools it usually visits during its summer recruitment season. Recognizing the limits of campus visits, the firm also made the games available online to first-year law students who were interested in applying to the firm.

"One of our primary objectives in looking at these tools was to broaden our pipeline of candidates," Snyder says. "It's very difficult under traditional law firm recruiting models to recruit people other than those you meet on campus. We were looking for something that would allow us to achieve a real and objective evaluation and go beyond campus interviews."

BIAS IN ITS MANY FORMS

AI's smart algorithms are trained to make inferences and conclusions based on massive data sets and machine learning. Like most technology, there is an impression of objectivity and precision because many believe technology must naturally be impartial.

That might have been true when evaluating contract language or reviewing a trove of emails, but the sophisticated analysis required in human resources or

continued on page 16

> "It's very difficult under traditional law firm recruiting models to recruit people other than those you meet on campus. We were looking for something that would allow us to achieve a real and objective evaluation and go beyond campus interviews.
>
> **DARIN SNYDER**
> ***O'Melveny & Myers LLP***

bank lending demands a higher standard than data magnitude. Data also must be stripped of demographic information that can lead to explicit bias in terms of race, age, disability, and gender, as well as information interwoven into the fabric of the data set such as zip codes and speech patterns.

Even the best efforts by technologists to remove implicit bias are no guarantee of success, say critics, and those biases can make AI tools particularly vulnerable to error. AI relies on certain suppositions — assumptions that can be built on incomplete data, bad coding, or inadequate models.

"The concept of 'garbage in, garbage out' generally applies to problem solving or policy making based on data analytics and AI," says Adrienne Fowler, a partner at Harris, Wiltshire & Grannis LLP and chair of its hiring committee. "Many data sets are inherently biased and will result in biased outcomes. But if you're vigilant, you can still work to reduce, or in some cases mitigate the impact of, bias in your data sets."

For example, Fowler notes, facial recognition data sets have more problems accurately identifying people of color than white people. If an AI program is based on large-scale data analysis and machine learning on facial recognition, then it's likely that data is biased. AI that offers insights and intelligence based on biased data will produce equally flawed outcomes.

"The games potentially allow you to achieve a certain degree of blindness in race, gender, and other factors that have led to discriminatory outcomes," says Fowler. "If you combine the games with good information about the skills you're looking for, try to test for analytical decision-making skills, and gather data that enables you to see whether or not the process you employ has a disproportionate impact on historically disadvantaged groups, then maybe you have a process that will work."

Despite the best intentions of everyone involved, it can be almost impossible in some cases to remove implicit bias because there are so many ways that bias can be infused, whether by accident, error, or fraud. However, some observers wonder if AI is being unfairly judged considering the alternative.

"AI always has this impossible standard to meet," says Dennis Kennedy, legal innovation consultant and author. "We keep moving the goal post on AI. I think at some point tech people could come back at critics and say, 'AI is bad compared to what?' It's not like humans are unbiased. In some cases, many of us would prefer to take our chances with AI."

If it is possible to introduce neutral, scientific, and reliable data into an algorithm, there are few disadvantages to relying on its outputs, argue some observers. The machine itself is not innately incorporating bias.

"Machines are dumb boxes until you start feeding them data," says Arriènne "Angel" M. Lezak, a shareholder with Polsinelli. "It's always going to come down to who is feeding the data and where are they getting it from. The human factor is going to be there because someone has to trigger the machine to make assumptions. The key is to have the data be as neutral as possible at the start."

Still, there are concerns that smart algorithms are disruptive because they are built on a weak structure. Some detractors say that cognitive games aren't likely to guarantee the results that law firms are looking for in recruiting or promotions.

"It's easy to assume that we're looking at causation here with cognitive empathy and high-performing partners," says Larry Richard, a lawyer and founder of LegalBrain, a consulting firm specializing in the psychology of lawyer behavior. "But does success cause the trait, or does the trait cause the success? The only way for any law firm to know for sure is to hire half the associates based on the games and [the other] half based on the old criteria and process. Six months out, the firm can look and see who is performing better, and then they'll know whether the games work."

> **"Many data sets are inherently biased and will result in biased outcomes. But if you're vigilant, you can still work to reduce, or in some cases mitigate the impact of, bias in your data sets.**
>
> **ADRIENNE FOWLER**
> ***Harris, Wiltshire & Grannis LLP***

ALGORITHMIC JUSTICE

Meanwhile, courts, prosecutors, and parole boards are turning to algorithms to objectively measure the likelihood that a defendant or parolee will commit more crimes, using that information to inform bail, sentencing, and early release decisions.

Gascón made a media splash when he announced his plans in San Francisco, in part because he hoped to clearly reduce racial bias. Under the new system, after a prosecutor makes an initial charging decision based on AI, he or she will review an unredacted version of the police report and other evidence that likely would reveal the race of the person. Only then will a final charging decision be made.

"This technology will reduce the threat that implicit bias poses to the purity of decisions which have serious ramifications for the accused," Gascón said in a statement. "That will help make our system of justice more fair and just."

In recent years, algorithm-driven risk assessments of whether defendants are likely to commit new crimes have been used widely in the criminal justice system and applied in a variety of cases, from determining bond to criminal sentencing.

Correctional Offender Management Profiling for Alternative Sanctions (COMPAS) is one such program. COMPAS incorporates 137 variables in its scoring algorithm. Research shows it has been moderately accurate in identifying recidivism, but African Americans are more often wrongly identified as a reoffending risk, according to ProPublica, the nonprofit newsroom that produces investigative reports.

"If computers could accurately predict which defendants were likely to commit new crimes, the criminal justice system could be fairer and more selective about who is incarcerated and for how long," noted ProPublica in its 2016 risk assessment reporting, "Machine Bias." "The trick, of course, is to make sure the computer gets it right. If it's wrong in one direction, a dangerous criminal could go free. If it's wrong in another direction, it could result in

someone unfairly receiving a harsher sentence or waiting longer for parole than is appropriate."

Challenging the validity of criminal justice data inputs is relatively simple in the United States, experts say, given that errors can be easily tracked to its long history of inequitable arrest and prosecutions due to unjust and racially biased laws.

"It's one of the fundamental problems of prediction in the criminal justice system," says Logan Koepke, senior policy analyst with Upturn, a nonprofit that advocates for equity in digital technology. "All predictions rely on historical data, and historical data reflects race discrimination and bias across the entire system. It's going to be very hard but necessary to find technical fixes that address the problem."

This is especially critical when AI exceeds its original productivity mission and is applied to the criminal justice system. Even minor flaws in the governing algorithms could exacerbate the problematic disparities that already mark the system.

"When you think about data for predictive policing or determining recidivism, most systems are going to use prior arrest records," says Koepke. "Arrests aren't the perfect measure of any action. If anything, arrest records document the practices and behaviors of police officers rather than being a true measure of crime."

MAINTAINING VIGILANCE

What experts suggest is that whoever uses AI to weed out implicit bias do so with the necessary due diligence at every stage of the process. Law firms should look for opportunities to audit the data sets, training parameters, and assumptions built into AI solutions and not just hope for the best.

"We are a society of data now," says Lezak. "People try to lean data one way or another all the time. It's as easy to manipulate data as much as it is anything else. It comes down to honesty, and a machine cannot guarantee honesty. What we need to do is teach people to look behind the data and at the source. You don't take anything at face value."

Others suggest that while the American business model uses proprietary algorithms and software that cannot be easily accessed, open-source products would be more effective in allowing for periodic assessments of the quality of the data and the impartiality of its sources.

Finally, LegalBrain's Richard observes that the biases built into human decisions are hard-wired: The known is preferred over the unknown, the familiar over the unfamiliar. While these biases have kept the human race alive, they also have established a pattern of fearing "the other." Machines have no legacy or context biases, but there is always the human connection.

"I don't think we can be completely rid of implicit bias," says Richard. "You really don't want to. It comes from a good place. What we need to do is understand and manage those situations where biases don't work, like in the workplace or the courtroom."

Sarah Kellogg is a regular contributor to Washington Lawyer.

Composite image, Getty Images/John M. Lund Photography Inc./DigitalVision; Darin Snyder, courtesy of O'Melveny & Myers LLP; Adrienne Fowler, courtesy of Harris, Wiltshire & Grannis LLP