iWitness

ARTIFICIAL INTELLIGENCE AND LEGAL ISSUES

RONALD HEDGES, GAIL GOTTEHRER, AND HON. JAMES C. FRANCIS IV

Ronald Hedges is a senior counsel at Dentons; Gail Gottehrer is the principal of the Law Office of Gail Gottehrer LLC; and the Hon. James C. Francis IV (Ret.) is an arbitrator, mediator, and special master with JAMS.

We live and practice law in a time of rapid—and sometimes confusing—technological change. Adoption of artificial intelligence (AI) is one. AI is no stranger to the practice of law. Clients use it to make business decisions; judges use it to assist their determinations, and experts in both civil and criminal litigation use it for their analyses. The output of AI can be relevant to issues that arise in litigation, and, because of its limitations, its use can invite legal challenge. Employment discrimination and criminal sentencing provide two examples.

AI refers to the development of computer systems that can mimic human decision-making and perform tasks that generally require human intelligence. AI uses algorithms, which are sets of rules that a computer can execute. Data are input into the algorithm, which applies those instructions and produces an output. Some artificial intelligence systems include algorithms that learn from data and improve automatically.

Mind-Sets in a Black Box

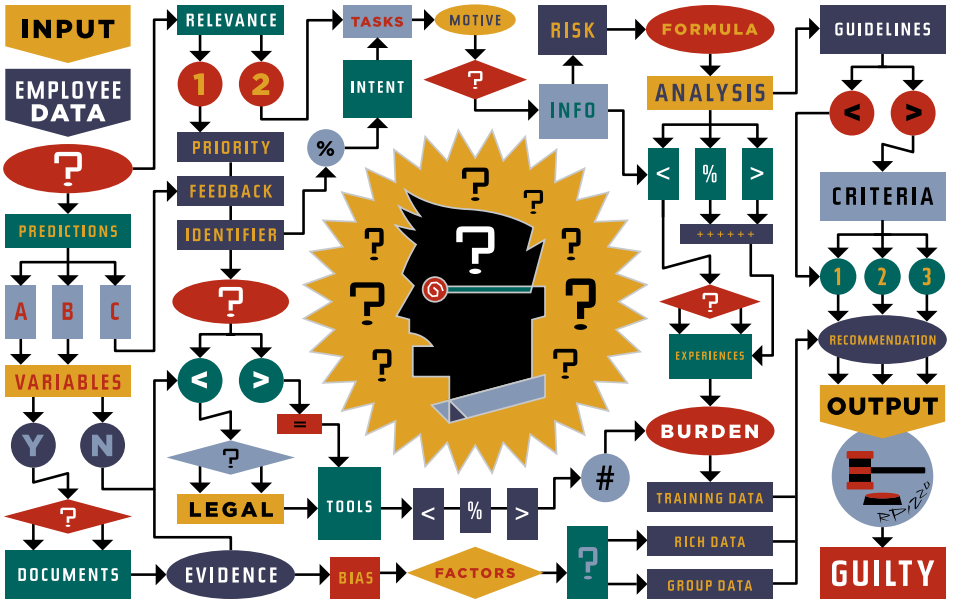
Many legal issues focus on the mind-set of the decision-maker, such as the intent of the manager who elects to terminate an employee, and on the process by which a person reaches a decision, such as the factors a judge considers when determining the appropriate sentence for a person convicted of a crime. The legal system has developed processes to examine how humans behave, but it is struggling to find an analytic framework to examine AI decision-making, which is often characterized by a lack of transparency into the processes by which it makes decisions and the biases of the programmers who create the algorithms on which it operates.

Critics observe that AI uses a decision process that is a "black box," meaning that it is based on algorithms that are so complex that the people who are affected by the decisions made by the systems cannot understand them and the government is unable to regulate them properly. The expanding use of AI in a wide range of industries, for significant decisions on everything from consumer credit limits to health care premiums, heightens concerns that the technology is effectively shielded from scrutiny by the complexity of the algorithms or trade secret considerations. The companies that design and use these algorithms consider them to be proprietary. Requests for disclosure of the algorithms and information about how they make their calculations are generally resisted on the grounds that these formulas are confidential business data that the companies are entitled to protect.

Attorneys are probably most familiar with AI's so-called "continuous active learning" in the context of technology-assisted review. Continuous active learning uses a machine-learning algorithm to find relevant electronic information within a large data set. It presents the reviewer with documents, ranked in order of priority, from those likely to be most relevant down to those likely to be less relevant. The reviewer provides input on the relevance of the documents presented by coding them, and based on that information, the system improves its understanding of which documents the reviewer considers relevant. The system continues to present documents to the reviewer and, through each step in this iterative process, "learns" from the feedback it receives and enhances its decision-making skills.

Continuous active learning systems also raise "black box" issues. These systems use algorithms to select the training data, rather than having humans select the data that are used to train the system. The role of the human in a continuous active learning system is limited to providing feedback to the system based on coding

Published in *Litigation,* Volume 47, Number 1, Fall 2020.



decisions. It does not include selecting documents to be used to train the system. As the training data are generated by algorithms, the data are unidentified and not preserved by the system, and they cannot be examined or challenged by litigators or regulators.

Bias

Bias is another fundamental concern associated with AI. While we may think of algorithms as simply being math, and therefore neutral, studies have shown that algorithms can be tainted by human bias.

Bias can be intentionally introduced into algorithms by the people who design them. Programmers can build bias into algorithms by relying on data that they know are biased against a certain racial or religious group or that reflect historical discrimination. In this way, biased programmers can skew the outputs of algorithms in a way that unlawfully discriminates against women or minorities. Programmers can also create biased algorithms by instructing an algorithm to consider, or give disproportionate weight to, factors that are proxies for sexism or racism.

Algorithms can also be affected by the implicit or unconscious bias of the people who program them. The design decisions of well-intentioned programmers can be influenced by their sociological background and experiences, which may lead them to rely on data that favor a certain group and disadvantage another group, without being aware of it. For example, using photos of White men to train facial recognition algorithms has resulted in those systems being proficient at recognizing White men but inaccurate when it comes to women and people of color. White male programmers may have trained those algorithms using photos of other White men without realizing that their training did not reflect society and would yield biased results.

Bias can also be an issue with continuous active learning systems. For example, if the algorithm uses skewed training data, it will create problems for the continuous active learning system and yield inaccurate and biased results. Similarly, if incomplete or insufficiently rich data are used to train the system, the results will reflect that deficiency and be biased.

If AI influences a decision as to who gets hired or promoted or receives a pay increase, it can be an instrument of unlawful employment discrimination. An algorithm for ranking candidates for promotion, for example, could explicitly include an identifier for race. More likely, it would include variables correlated with race, such as whether the candidate has graduated from high school or has an arrest record or lives in a certain ZIP code. If the AI is an active learning tool—one that identifies candidates most like those who have proven successful in the past—then, if the employer has historically favored employees of a particular race, the algorithm will do the same.

Addressing Biased Artificial Intelligence

How, then, might a legal framework like Title VII of the Civil Rights Act of 1964 address bias that arises from the use of AI?

Employment discrimination cases under Title VII generally fall into two categories: disparate treatment and disparate impact. In disparate treatment cases, the plaintiff must show that the employer intended to discriminate on the basis of some prohibited characteristic. But intent is difficult to demonstrate when the decision-maker is not a person, but an algorithm.

> A defendant may not know what information is being taken into consideration by an algorithm in making a risk assessment.

Suppose a Black employee alleges discrimination because he was denied a promotion that went to an arguably less qualified White worker. This plaintiff would have met the minimal burden of providing facts from which an inference of discrimination could be drawn. The employer, however, could then present evidence

Illustration by Robert Pizzo

that it based its decision on a promotion algorithm. At that point, it is the plaintiff's burden to demonstrate that, notwithstanding the seemingly race-neutral basis for its decision, the employer's real motive was discriminatory.

The plaintiff might attempt to show that the employer created the algorithm with the intention of favoring White employees, but few employers develop their own AI. More likely, the employer has licensed the AI from a third party, and the plaintiff might try to demonstrate that the employer knew that relying on the tool would have a discriminatory effect. Finally, the plaintiff might strive to demonstrate that the output of the AI tool was dependent on input from the employer and that the employer knowingly provided biased data.

In disparate impact cases, the employer uses a facially neutral method for allocating employment benefits, which is claimed to have the effect of discriminating on the basis of a proscribed characteristic such as race. A promotion algorithm may be such a neutral tool. To challenge it, the plaintiff need not show discriminatory animus but would likely rely on statistical analysis demonstrating that use of the AI resulted in White candidates being favored over Black candidates for promotion, after controlling for factors other than race.

Once the plaintiff has demonstrated that the algorithm has a disparate impact, the employer is entitled to demonstrate that it is nevertheless performing a legitimate function—here, the selection of qualified applicants for promotion.

If the employer satisfies this obligation, the burden shifts back to the plaintiff to show that other selection devices, without a similarly disparate effect on Black candidates, would also serve the employer's legitimate interest. Thus, the plaintiff must not only demonstrate the differential impact of the algorithm, but also identify an alternative, nondiscriminatory selection method that meets the employer's needs.

An employment discrimination plaintiff, then, can challenge an employer's use of biased AI, but that plaintiff will confront significant hurdles, whether proceeding on a disparate treatment or disparate impact theory.

Artificial Intelligence in Criminal Proceedings

AI has been used in criminal proceedings for some time. Decisions involving, for example, bail, sentencing, and parole may be influenced by "recommendations" made by pretrial or probation departments that are based at least in part on the output of AI. These recommendations—and the judicial decisions that result from them—can give rise to due process issues.

The use of AI in the criminal context presents several basic questions. First, should predictions of future criminal behavior be considered at all in pretrial, sentencing, or parole decisions? Second, when those predictions are made by AI, are they subject to the types of bias discussed above? Yet, even if there is consensus that predictive AI is appropriate in the criminal context and that problems of systemic bias can be addressed, a fundamental question of due process remains.

We have one decision on point: *State v. Loomis*, 881 N.W.2d 749 (Wis. 2016), *cert. denied*, 137 S. Ct. 2290 (2017). *Loomis* was a challenge to the use of the COMPAS risk assessment tool. The defendant had pleaded guilty to various minor offenses arising out of a drive-by shooting. In preparation for sentencing, the State presented an estimate of the risk of recidivism based on COMPAS's analysis of an interview with the defendant and information from his criminal history. The defendant was sentenced to a term of imprisonment and challenged the use of the tool on due process grounds, arguing that he was denied his due process rights to be sentenced on an individualized basis and on accurate information. In particular, Loomis asserted that because the algorithm was a black box, he was unable to challenge how it came to its conclusions. The trial court denied Loomis relief, and the Wisconsin Supreme Court affirmed. Although it raised questions about the tool, the Wisconsin Supreme Court cautiously approved its use, noting that the assessment was only one factor in a judge's sentencing decision, which the judge could disregard. Moreover, the court observed that "due process implications compel us to caution circuit courts that 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."

Where does *Loomis* leave us? First, a defendant may not know what information is being taken into consideration by an algorithm in making a risk assessment. Second, a defendant almost certainly does not know how the algorithm weighs various data to produce an assessment. While a criminal defendant might seek discovery of the algorithm, its proprietary nature would be a major obstacle.

Of course, this should all be put into context. Federal and state sentencing guidelines establish criteria that judges follow in making sentencing decisions, and those guidelines often rely on calculations that assign values to various offender characteristics. Nevertheless, these guidelines are transparent about the formula they rely on for reaching an ultimate recommendation. AI, by contrast, is opaque.

The use of AI in litigation introduces complexities and invites considering whether its shortcomings may be grounds for litigating decisions based on its output. There is much more that can and will be said about artificial intelligence including, among other things, how it might be challenged in other contexts, whether counsel must become competent in AI technologies, and when it is appropriate to take discovery into how an algorithm "works" and delivers its outcome. ■