ARTIFICIAL INTELLIGENCE & IMPLICIT BIAS

WITH GREAT POWER COMES GREAT RESPONSIBILITY

ADDRESSING THE BIASES INHERENT IN THE DATASETS THAT DRIVE AI APPLICATIONS AND THEIR ALGORITHMS.

BY KIM P. NAYYER, MARCELO RODRIGUEZ & SARAH SUTHERLAND
We are all familiar with news headlines such as Fast Company’s “Biased AI is a Threat to Civil Liberties,” Wired’s “Algorithms Should’ve Made Courts More Fair. What Went Wrong?,” and the Washington Post’s “Algorithms Were Supposed to Make Virginia Judges Fairer. What Happened Was Far More Complicated.” Systems, procedures, and workflows using artificial intelligence (AI) are becoming more common at every level of our judicial system. Given their decisive significance and the life-altering decisions they often make, it is important that all of us understand what is really taking place when these automated systems become a standard and, eventually, indispensable part of our legal system.

Implicit bias can derive from the natural and normal subconscious or cognitive biases that invariably affect us all, as well as from long-term societal conditions and practices. In his booklet, “Invisible Influences: Basic Facts on Implicit Bias, Stereotype Threat, and Privilege,” presented during the 2019 American Association of Law Libraries (AALL) Diversity & Inclusion Symposium, Russell A. McClain identifies institutional bias, systemic bias, and subconscious bias as sources or vehicles of implicit bias in everyday decision-making.

Due to its pervasive and hidden nature, implicit bias forms what might be thought of as an invisible lens distorting what appears at every stage of the selection and curation of datasets, as well as magnifying imbalances in the perspectives of every person involved in creating these datasets. Moreover, until recently, the size, scope, and content of datasets used in the training of machine-learning algorithms were not well understood. Arguably, little attention was paid to the quantity, quality, and representativeness of the training data. It is recognized now that training data is likely replete with institutional, systemic, and subconscious biases. Developers, most likely inadvertently, have been creating AI tools that have been learning from data that magnify systemic biases and distort outcomes.

When it comes to the use and application of AI in our legal system, the unseen implicit bias in datasets reflected in machine learning and automated decisions may accentuate patterns of discrimination. Automated systems and datasets do not tend to take into account the complexity that goes into judicial decisions, the concepts of fairness and due process, or the social context. Implicit bias in our judicial system is not a new phenomenon. However, the cognitive phenomenon of automation bias—the normal tendency to trust a computer to make decisions—is at play. The explosion of AI-based legal data products has brought these issues to the fore in explicit and dramatic ways: ignorance and idle consumption are not viable options.

**What Should We Do?**

Some academics advocate banning AI-based products or features from legal systems, while others advocate for “rigorous regulatory guidance” that will provide guidelines and a framework for all actors utilizing AI-based products. In contrast, some praise the transparency and potential accountability in datasets, which may subsequently be “fixed,” thus alleviating some of the concerns regarding bias embedded in the data.

In 2019, the American Bar Association (ABA) approved the following resolution: “The American Bar Association urges courts and lawyers to address the emerging ethical and legal issues related to the usage of artificial intelligence (AI) in the practice of law including: (1) bias, explainability, and transparency of automated decisions made by AI; (2) ethical and beneficial usage of AI; and (3) controls and oversight of AI and the vendors that provide AI.” The resolution encourages attorneys and law firms to not assume the effectiveness of AI products. Instead, attorneys are encouraged to ask vendors about the ins and outs of datasets and their application before acquiring AI products, and to conduct periodic audits into the datasets and their implementations. The ABA encourages its members to conduct themselves under the ethical duties of: (1) competence (and diligence); (2) communication; (3) confidentiality; and (4) supervision when operating AI products for litigation.

**AI in the Justice System**

Part of the problem with the application of AI in the justice system is that the structure of legal data itself makes it difficult to assess it and base statistics and predictions on it. These problems arise from multiple sources. The first major source of error is due to the data that is available for analysis. This data does not usually reflect the full set of similar situations, as it frequently excludes matters that settle out of court or where a judge issued only an oral opinion. The missing data tends to be the most common outcomes, because there are no unusual issues to litigate.

For valid outcomes, statistical analysis typically relies on data being in a normal distribution—that is, a pattern where the most common results are in the middle, and less common data points are more extreme. This pattern is found in many naturally occurring
situations, and it allows for comparisons of averages and identification of situations where it is not likely for differences in outcomes to be random. In legal data, the middle of the normal distribution is often absent from the set. So, while the calculations can be run against these sets and will give results, they may not be valid. If more statistical analysis were done on datasets that reflect all outcomes for everyone, it would allow for quantitative comparisons based on different variables and make bias more easily quantified. In the absence of this analysis, it’s difficult to assess whether the data is appropriate for use in AI systems or to identify appropriate AI-driven research questions.

A second major source of error is not due to mathematical or statistical errors. Instead, these errors arise because there are undesired patterns in existing data, and AI systems reflect and may even magnify them. AI systems frequently don’t display why certain outcomes are suggested, so this can make biased recommendations difficult to identify. To counter this issue, it is common to audit results based on outcomes to see how closely they conform to what we expect them to be.

Adjusting for or mitigating errors derived from biased or incomplete data in AI systems is complicated. It may require extensive testing and interventions such as addition or removal of particular data elements or adjusting recommendations to give statistically balanced results. That said, without the ability to break down the results from an AI black box, it is difficult to know whether the basis for a decision is fair or something that would be admissible in court. To improve this requires identifying the sources of bias in historical data and deciding what is fair in some of the most difficult decisions we make.

Is It All About the Lawyers?
We know that uncertainty exists in our ability to assess fairness and reliability. Where do these challenges leave us, the legal information professionals, the practitioners using AI-driven tools? What are the ethical implications for attorneys if functions such as predictive analysis, decision support, or legal research and analysis incorporate AI that may be subject to implicit bias? If available datasets are flawed, in that they are not representative and are too small to yield high-quality machine learning outcomes, and if our goals include assisting or replicating complex human reasoning that we know can yield multiple valid outcomes, and that same reasoning also reflects inherent biases of society, are we not then creating tools that are in turn perpetuating bias? Should developers continue to experiment and innovate, given the limitations of available datasets?

Attorneys clearly bear a range of strict professional ethical and non-discrimination obligations in licensing requirements and professional oaths. In our view, attorneys, or those accountable at the highest level, must also bear some measure of responsibility to understand and assess the realities of the outputs of current AI systems in use in their practices, even as the inputs and processes may remain black boxes.

Information Professionals Have Something to Say
We venture that such ethical concerns and responsibilities extend to legal information professionals who may be involved in acquiring, operating, or even building such AI tools and systems. Beyond this, our community of information professionals operates with its own ethical canons. Whereas the professional associations or bodies responsible for accreditation of our educational programs may not yet have offered specific ethics opinions on the use of AI in law libraries or on our role in the larger provision of legal professional service delivery, existing ethics expectations remain. We have a duty of competence and a duty to supervise, among other professional responsibilities, that may intersect with our work with AI tools that are subject to the impact of implicit bias in unknown measure. We suggest it may be time to add to these duties an expectation to prevent, minimize, or qualify AI applications to enable libraries to maintain high standards of ethical responsibility.

It is essential that innovators who create and use AI applications in law libraries be aware of the wide range of ethical implications that may arise as a result of embedded or entangled implicit bias. The ability to understand these impacts of AI systems—or, at a minimum, the realities of their current limitations, their outputs, and their potential impact on populations—is at least as important as our other professional responsibilities. Workers in law libraries cannot make ethical decisions and demonstrate conduct that is as free as possible of bias unless they recognize and are responsive to the challenges that are inherent in the datasets and algorithms driving AI applications.