THOMSON REUTERS
Westlaw®

'Your AI is Racist.' An accusation that is probably more accurate than you think

By Brad Goldstein, *ProCrysAI LLC*, and Tod Northman, Esq., *Tucker Ellis LLP*

MARCH 19, 2021

Your AI is racist. It is probably also sexist, ageist, and elitist. The existence of bias in artificial intelligence (AI) systems has gained much attention as of late.

From mortgage approval algorithms to facial recognition systems to medical diagnostic support systems, the popular press has even begun to take note. So, too, should any organization that develops, sells, or utilizes AI and ML (machine learning) systems.

> If bias goes into the system, inevitably bias will come out of the system.

The threats of liability can be subtle but are very real at all points in the healthcare/medicine delivery system. Care must be taken to understand what bias is, where it comes from, and how to mitigate its potential negative effects.

Only in this way can a systematic approach to preparing for the legal aspects of bias be initiated.

WHAT IS BIAS IN ARTIFICIAL INTELLIGENCE?

We need to start with a reasonable definition of bias as it pertains to AI/ML. There are multiple forms of bias that can be introduced into an AI system. Dr. Amol S. Navathe breaks bias in healthcare and medicine into two distinct types: statistical bias and social bias.[1]

Statistical bias is created when an algorithm produces results that are not representative of the true population.

For example, radiology AI to be used for lung cancer detection and diagnosis would be biased if it was trained on data skewed toward individuals with health insurance coverage that includes screening for past or present smokers.

Individuals whose insurance does not include this benefit and those without health insurance would be underrepresented in the data. The sample would not be reflective of the actual population of past and present smokers.

Social bias is more akin to what we usually think of when we hear the term bias: a particular group, or population, is being treated in an unequal fashion within the system.

In healthcare, this can lead directly to less than optimal outcomes for patients, their families, and their communities.

A salient example of social bias — and one that received much attention — was seen in the performance of AI-based facial recognition systems used in law enforcement in 2019.

Asians and Blacks were misidentified by these systems up to 100 times more than white men.[2]

In another example, Sam Bowman and others at New York University found racial bias and cultural stereotyping historically embedded and reinforced in internet search engines.[3]

The acknowledgement and recognition of the import of the presence of bias cannot be overstated. AI systems are not used in isolation.

The very real-world settings within which these AI systems are utilized make it highly possible that they will cause some degree of unintended harm. Diligence is an obligation.

The performance of AI and ML is circumscribed by the quality of data used for training. Underrepresentation of communities and populations, statistical outliers, and simply missing data, can all negatively affect the generalizability of an AI algorithm.

If bias goes into the system, inevitably bias will come out of the system. In fact, the bias can be amplified by its own perceived success. As the AI makes recommendations — even bad ones that are accepted by the end user — those recommendations are reinforced and perpetuated.

Because the data contained within any given AI/ML training set was created by humans, it will contain evidence of the biases that are entrenched in our organizations, institutions, and society in general. Bias will always produce suboptimal outcomes.

This is especially true in algorithms created to perform a human function, such as medical diagnoses. The data that the system was trained on was generated by humans, in this case largely by physicians.

The algorithm will learn any assumptions or inferences that the all-too-human physicians inadvertently input into the data.

It might seem highly unlikely that a practitioner would accept a recommendation based on faulty machine learning; however, frequently the assumption is held that data is just a set of facts and therefore bias free.

There is also the problem of "confirmation bias," a form of statistical bias in which information that supports, for instance, a physician's assumptions and beliefs about an outcome, will be accepted as valid. In these ways, bias can be perpetuated and overlooked in an AI system.

WHAT CAN BE DONE?

We use AI because it is supposed to perform better than humans at its appointed task. How can we expect AI to improve patient outcomes if our prejudices, inferences, and assumptions are codified directly into the system?

Bias must be accounted for and mitigated in existing systems and prevented and tested for in systems under development.

It is far from all bad news when it comes to bias in AI/ML. Bias can be mitigated and reduced in both existing and new AI systems. In fact, in some instances, AI itself can be used to accomplish these goals.

How can we expect AI to improve patient outcomes if our prejudices, inferences, and assumptions are codified directly into the system?

AI can be used to test existing algorithms on diverse population sets. The resulting outputs would be compared in order to isolate any discrepancies revealed. The algorithm could then be adjusted to provide appropriate non-biased outputs.

Decision support AI can be developed to specifically look for biased clinical decision making on a real-time basis. These algorithms could alert physicians to a discrepancy in their clinical practice from that of the expected non-biased norm.

The algorithm could pick up on race — or gender-skewed keywords that were put into a patient's electronic health record (EHR).

For instance, a female patient might be characterized as "hysterical," while a male patient reporting the same condition might be characterized as "agitated."

Not only is this difference in characterization sexist, but, if this data is allowed to be used to train the AI, it will introduce inaccuracies into the system.

The AI could flag this issue for the physician instantly, thus giving them the opportunity to correct the error and improve the quality of data entering the EHR.

As in most complex professions, a set of best practices needs to evolve in order to ensure the highest quality outcomes. This is especially true in medical AI where patient safety can literally be a life or death situation.

THE EMERGENCE OF BEST PRACTICES

For systems under development, as well as those being adopted throughout an institution, the opportunity exists to address the issue of bias head-on. Much can be done to eliminate bias in how the data is initially collected.

Careful consideration of appropriate data collection paradigms coupled with an understanding of the organization's data collection infrastructure are necessary to succeed.

Below is a small but useful subset of best practices that can make a big difference:

- Data should be collected on an everyone-through-the-door basis and throughout the entire patient encounter. This style of comprehensive data collection decreases the opportunity for bias to be introduced into the learning data set.

- Data needs to be representative with as much input from the patient perspective as possible and not heavily weighted by the clinical assumptions and decisions of healthcare providers.

- Data-gathering quality systems need to be put into place. Constant monitoring and reviewing practices aid in preventing bias from entering the data at all stages of collection. Guidelines such as the PROBAST (Prediction model Risk Of Bias ASsessment Tool) tool are available and periodically updated to help organizations implement adequate mitigation strategies.

- Algorithmic stewardship, as outlined by Zaid Obermeyer, MD, puts the onus on an organization to be aware of and continually monitor all AI/ML systems deployed and in use.[4]

By implementing best practices at every phase of the AI/ML supply chain, an organization can work toward bias-free AI.

Quality systems, reflective of the current regulatory landscape, will need to be developed to mitigate bias at every point of algorithm creation. Early adoption of appropriate best practices will be key to maximizing their benefit and allowing for adaptation and growth.

WHAT ARE THE LEGAL IMPLICATIONS?

As an example, racial disparities presenting as negative patient outcomes can be a probable result of a biased medical AI. This is where accusations, such as "Your AI is Racist"

become very real, very problematic, and quite possibly, as we have seen above, accurate.

Patients and/or their families could seek legal remedies. Standard product liability principles, with extrapolations to fit the particular AI/ML system, will be employed to find fault.

It is better for a client to understand up front that bias will exist in these systems and to have a specific and comprehensive business, regulatory, and legal plan in place.

Any plan will need to have ongoing review. Adaptations to adjust, in an agile fashion, as the AI landscape quickly evolves will be necessary.

By the time there is a crisis, it will be too late for the ill prepared to avoid the expense of litigation and the probable harm done to the reputation of the organization. Due to a prevalent misunderstanding and mistrust of AI, public opinion will be harsh.

Being able to demonstrate extreme diligence by taking all currently available steps to mitigate and correct bias is a necessity, not a budgetary nicety.

The adoption of best practices and proactive mitigation strategies in the near term is essential. AI is an ever-evolving field. The implementation of these mitigation strategies is not where it ends for an organization, but where it all begins.

Given this, it is inevitable that legal action will be taken — with or without merit — by patients, their families, or third-party organizations acting on their behalf. Documenting an organization's bias mitigation strategies will be paramount to any defense.

Not only is an internal team of AI "stewards" necessary, but it is imperative that organizations work with external legal and business/regulatory experts.

An external support team can provide the insight necessary to stay one step ahead of changes on the best practices/regulatory/legal playing field.

There will be the tendency, as mentioned above, for AI in medicine and healthcare to be judged harshly. AI is frequently demonized.

Bias is only one of the factors that plays into this perception. Organizations involved in creating or utilizing AI may ultimately be subject to "strict liability" or, at a minimum, product liability claims.[5]

The basis of these cases will be that bias is inevitable, and therefore AI is de-facto "unreasonably dangerous." Without well thought-out and fully implemented best practices, it will be impossible to prove otherwise.

Being able to demonstrate extreme diligence by taking all currently available steps to mitigate and correct bias is a necessity, not a budgetary nicety. Organizations must own their responsibility to provide safe and effective products and services to their patient base.

Awareness of the presence of bias in AI/ML is growing. Appropriately, a commitment to ensuring the bias-free development and application of AI in healthcare and medicine is gaining traction.

Anecdotal evidence of this growth in awareness is the contrast between the October 2019 and the December 2020 (virtual) HIMMS (Healthcare Information and Management Systems Society) Machine Learning & AI for Healthcare conferences, both of which were attended by co-author Brad Goldstein.

Bias was very much on the minds of the organizers and participants.

By 2020, two leaders in the field of bias in AI — Zaid Obermeyer, MD of the Berkeley School of Public Health and Kadija Ferryman of the New York University Tandon School of Engineering — were speaking to a very receptive audience.

Although issues of liability arising from bias were not directly addressed, its time has come.

Notes

[1] Navathe AS, Parikh RB, Teeple S. (Dec. 24/31, 2019) Addressing Bias in Artificial Intelligence in Health Care. JAMA 2019; 322(24):2377-2378. doi:10.1001/jama.2019.18058

[2] *Buolamwini*, Joy. "Gender shades: Intersectional accuracy disparities in commercial gender classification". *Conference on Fairness, Accountability and Transparency. (2018) 81: 77-91*

[3] New York University. (Oct. 1, 2020). Researchers Spot Origins of Stereotyping in AI Language Technologies.[Press release]. https://bit.ly/3rY2QaF

[4] Butte AJ, Eaneff S, Obermeyer Z. (Oct. 13, 2020) The Case for Algorithmic Stewardship for Artificial Intelligence and Machine Learning Technologies. JAMA 2020; 324(14):1397-1398. doi:10.1001/jama.2020.9371

[5] See discussion at Abbott, Ryan Benjamin, The Reasonable Computer: Disrupting the Paradigm of Tort Liability (Nov. 29, 2016). George Washington Law Review, Vol. 86, No. 1, 2018, Available at SSRN: https://bit.ly/3bTCLE6 or https://bit.ly/3vAagmt

This article was published on Westlaw Today on March 19, 2021.

ABOUT THE AUTHORS

Brad Goldstein (L), principal at **ProCrysAI LLC** in Beachwood, Ohio, has an academic background in electrical/computer engineering and health care. He is an entrepreneur and consultant who has spent over 20 years managing the development and deployment of advanced technologies. He can be reached at bradg@ProCrysAI.com. **Tod Northman** (R) is a partner in **Tucker Ellis LLP**'s Cleveland office whose practice focuses on corporate and emerging tech law and transactions. Northman is co-chair of the firm's Autonomous Vehicles & Artificial Intelligence Technologies Group. He can be reached at tod.northman@tuckerellis.com.

Thomson Reuters develops and delivers intelligent information and solutions for professionals, connecting and empowering global markets. We enable professionals to make the decisions that matter most, all powered by the world's most trusted news organization.