The Legal Risks Of Bias In Artificial Intelligence

By **Patrick Huston and Lourdes Fuentes-Slater**

Artificial intelligence algorithms are designed by humans and work off of data.
Both humans and historical data can be biased. It is, therefore, easy to see how bias can enter into an algorithm.

AI has now become commonplace in the digital world, so it is important to have a proactive approach to dealing with the consequences of AI, including algorithm bias. However, we should preface a discussion on AI bias by noting that bias is a byproduct — or consequence — of AI. Some consequences are intentional. Others are not.

While the word "bias" may carry a negative connotation, not all biases are bad. Bias is a tendency toward something. Some biases are positive and helpful, like a bias for healthy food or exercise, or avoiding dangerous activities or places.

Also, some algorithms are purposefully designed to target specific audiences or have the tendency toward something. For example, some are designed to be biased toward a commercial or social target. They may be geared to certain people's choices and interests, such as dating-site algorithms, algorithms targeting the rich or the poor for commercial reasons, algorithms targeting women versus men, algorithms aimed at certain demographics, etc.

Therefore, the goal is not the eradication of all bias. Even if that was feasible, it is neither legally necessary nor commercially sound. Rather, our goal is to understand the legal implications of some algorithms' propensity to have illegal or harmful impacts due to negative biases so that we can take reasonable steps to mitigate those particular biases.

What is algorithm bias?

Algorithm bias is a phenomenon that occurs when an algorithm produces results that are systematically prejudiced due to assumptions in the machine learning process.[1] Bias in AI systems has two major sources: the data sets on which models are trained, and the design of the models themselves.[2]

The first source is the datasets on which algorithms are trained. This data carries human biases with it. For example, biases in the judicial system affect who gets charged and sentenced for crimes. If that data is then used to predict who is likely to commit crimes, then those biases will carry over.

A second major source of bias results from how decision-making models are designed. For example, if a teacher's ability is evaluated based on test scores, then other aspects of performance, such as taking on children with learning differences or emotional problems, could fail to register, or even unfairly penalize them.

What are the risks with algorithm bias?

One of the risks with algorithm bias is that it will result in unintentional discrimination. For example, the Fair Housing Act "prohibits housing-related discrimination on the basis of race, color, religion, sex, disability, familial status, and national origin." A recent study, however, raised valid concerns that reliance on algorithms in the housing industry could result in housing discrimination.[3]

AI bias has even impacted COVID-19 mortality predictions due to a myriad of sources for potential bias. Biological associations between race, gender, age and these biomarkers could lead to biased estimates that don't represent mortality risk. Unmeasured behavioral characteristics can lead to biases, too.[4][5]

It is important to recognize that the biggest challenge with AI bias is that, unlike human bias, it can grow exponentially. A small bit of bias in data can have a huge ripple effect. Herein lies the problem. In recognition of this reality, the American Bar Association last year issued a resolution urging courts and lawyers to address the emerging ethical and legal issues related to the use of AI.

The reality is that we can easily envision the argument that the harmful impact of AI algorithm bias should have been reasonably known and that a failure to account for them would be negligent. Given that the discussion over the presence of AI bias is raging already, what was known or reasonably should have been known is a valid debate. The consequences of algorithm bias are ripe to become the next wave of class actions.

Indeed, those that should be concerned with AI bias are taking notice. For one, the U.S. Department of Defense — one of the world's largest AI users — is taking steps to ensure the legal and ethical use of AI and other emerging technologies. It recently announced that one of its five AI principles is specifically focused on taking "deliberate steps to minimize unintended bias in AI capabilities."[6]

What steps can be taken to ensure fairness in algorithm functions?

Legally "protected classes" in the United States generally include age, race, gender, religion, color, national origin, disability and ethnicity. As a general matter, any algorithm that can have a disparate impact on a protected class should be reviewed carefully at the design and development stage and then audited regularly. Therefore, companies should take proactive steps to ensure that AI used for business purposes is not prone to algorithm bias on the protected classes. Failure to do so may result in legal actions against them.

There are steps being taken toward that goal:

1. Human in the Loop or Human-Machine Teaming

The greatest potential for AI will come from teaming humans and machines in ways that leverage the respective strengths of both. Machines perform some tasks better than humans, such as: rapid computations, analyzing mass data, and performing boring, repetitive tasks. But there are traits where humans generally have the edge, such as leadership, judgment and common sense. Human-machine teaming, or HMT, merges the best attributes of humans and machines in order to optimize outcomes.

Human in the loop, or HITL, should be involved in the design, utilization, selection and back-end audit of an AI system, particularly where unmonitored automation carries significant risks. The HITL should (1) understand the reasoning behind the decision and the factors underlying the decision and (2) identify a natural or legal person (i.e., not the AI systems algorithm itself) who would bear liability and thus be accountable for violations against transparency.

Where a human is impacted by the utilization of an AI system, the HITL should be put in a position where they can (1) become aware that the human has been subject to a decision made by an AI algorithm and (2) determine whether the HITL would be willing to allow the algorithm to continue to process personal information pertaining to and to make decisions affecting the human.

2. Regulation

AI regulation is still at its nascent stages; however, we have seen a clear move toward regulation both here and abroad. Last year, Congress introduced legislation that would regulate certain aspects of AI.[7]

The Algorithmic Accountability Act of 2019 would require large companies to audit their algorithms for potential bias and discrimination, and to submit impact assessments to Federal Trade Commission officials. The reports would have to address the accuracy, fairness, bias, discrimination, privacy and security issues of any high-risk systems being used. The reports would also have to advise the FTC on how the system was developed and the data it uses.

As currently written, the act targets large companies by limiting its application to those with more than $50 million in gross annual revenue or those possessing personal information on more than 1 million consumers or consumer devices.

There is also the Commercial Facial Recognition Act of 2019, introduced in March, which would generally ban the commercial use of facial recognition technology to "identify or track an end user" without obtaining their consent.

It would also prohibit: (1) using "the facial recognition technology to discriminate against an end user"; (2) repurposing "facial recognition data for a purpose that is different from those presented to the end user"; and (3) sharing "the facial recognition data with an unaffiliated third party without affirmative consent."

Under the bill, with some limited exceptions, facial recognition technology that is available as an online service must also be made available for independent third-party testing "for accuracy and bias."[8]

At the state level, we are also seeing the same trend. In December 2017, the New York City Council passed Local Law 49, the first law in the country designed to address algorithmic bias and discrimination occurring as a result of algorithms used by city agencies. In November 2019, the task force issued its final report. An algorithms management and policy officer was appointed to serve as a centralized resource on algorithm policy and develop guidelines and best practices to assist city agencies in their use of algorithms. The new officer will also be responsible for ensuring that algorithms used by the city to deliver services promote equity, fairness and accountability.[9]

Even private companies are taking a stance. Several large technology companies have taken a surprising stance in favor of federal regulations to combat AI bias. Last year, for example, Microsoft's president said, "we need legislation that will put impartial testing groups like Consumer Reports and their counterparts in a position where they can test facial recognition services for accuracy and unfair bias."

Some speculate that this stance is being driven by a preference for a single overarching federal statute over the alternative — a patchwork of rules in different states that would make compliance difficult for companies that operate nationally or internationally.[10]

Lastly, but perhaps leading the way in this area, is the European Commission, much like it did with privacy regulation under the General Data Protection Regulation, it aims to play an early and key role in AI governance. Earlier this year, the EC published its "White Paper on Artificial Intelligence - A European approach to excellence and trust," indicating that an entire regulatory framework around AI development and application will be forthcoming.[11]

3. Auditing Algorithms

Algorithmic auditing is a collection of techniques for testing whether an AI system has biases. The idea behind an algorithm audit is to examine the inputs, outputs and outcomes in a scientific way to ensure they are working as intended. We must invest, develop and cultivate these mitigation tools.[12]

A place to start would be with IBM's AI Fairness 360 toolkit, an open-source library to help detect and remove bias in machine learning models. Its Python package, for example, includes a comprehensive set of metrics for datasets and models to test for biases, explanations for these metrics, and algorithms to mitigate bias in datasets and models.[13]

Auditing algorithms is a proactive way to avoid illegal algorithm bias and is something lawyers should be aware of.[14] Private companies and regulators are exploring this type of auditing as a possible solution to stop illegal bias. Indeed, an Accenture survey proposed three new likely categories of human jobs — trainers, explainers and sustainers.[15]

Trainers teach AI systems how to perform, process data and behave. Explainers improve the transparency of the AI inner workings. Sustainers will do audits to ensure that AI is fair, safe and responsible. Auditors would examine data sources , the choice of analytical tools, and the way in which results are interpreted to ensure that, as far as possible, built-in biases are eliminated.[16]

4. Ethicists

We also need ethicists involved in the development of AI models, in addition to data scientists and technologists.[17] Military AI programs have come under intense scrutiny, but AI ethics experts have helped temper the concerns.

For example, the U.S. Army's AI Task Force appointed the head of the West Point Philosophy Department to serve as its ethics expert. Large technology companies such as Microsoft and Google have followed suit and appointed ethicists to their AI teams. Their ethical perspectives often provide a view that AI technologists don't see.[18]

5. Explainable AI

Deployers of AI tools should provide appropriate transparency of the types of data used by the algorithm to make the decision and the impact of the decision on the individual — otherwise known as "explainable AI." Explainable AI is also one of the Pentagon's new AI principles. The Pentagon calls it "traceable AI" and states that "AI capabilities will be developed and deployed such that relevant personnel possess an appropriate understanding of the technology, development processes and operational methods applicable to AI capabilities, including with transparent and auditable methodologies, data sources and design procedures, and documentation."[19]

Conclusion

AI is so common that the risk of algorithm bias is already all around us. We should avoid illegal bias wherever we can and carefully manage the remaining bias in order to minimize the risk of inevitable algorithm bias lawsuits. Some ways to address this AI risk include: human-machine teaming, governing AI with realistic and consistent regulations, leveraging AI auditors and ethicists, and ensuring AI remains "explainable." These steps, when done in concert, will allow today's users to get the most out of AI, while minimizing AI's inherent risks.

Brig. Gen. Patrick Huston is assistant judge advocate general for military law and operations at the Pentagon.

Lourdes Fuentes-Slater is founder and CEO at Karta Legal LLC.

The opinions expressed are those of the author(s) and do not necessarily reflect the views of the firm, its clients, or Portfolio Media Inc., or any of its or their respective affiliates. This article is for general information purposes and is not intended to be and should not be taken as legal advice.

[1] Rouse, Margaret. "Machine learning bias algorithm bias or AL bias." TechTarget (Blog), Oct. 2018, available at https://searchenterpriseai.techtarget.com/definition/machine-learning-bias-algorithm-bias-or-AI-bias.

[2] Satell, Greg, Sutton, Josh. "We Need Al That Is Explainable, Auditable, and Transparent." Harvard Business Review, Oct. 28, 2019, available at https://hbr.org/2019/10/we-need-ai-that-is-explainable-auditable-and-transparent.

[3] Villasenor, John, Foggo, Virginia. "Why a proposed HUD rule could worsen algorithm-driven housing discrimination." Tech Tank, Brookings, Apr. 16, 2020, available at https://www.brookings.edu/blog/techtank/2020/04/16/why-a-proposed-hud-rule-could-worsen-algorithm-driven-housing-discrimination/?utm_campaign=Center%20for%20Technology%20Innovation&utm_source=hs_email&utm_medium=email&utm_content=86618964.

[4] Engler, Alex. "A guide to healthy skepticism of artificial intelligence and coronavirus," Al Governance, The Brookings Institution's Artificial Intelligence and Emerging Technology (AIET) Initiative, Brookings, Apr. 2, 2020, available at https://www-brookings-edu.cdn.ampproject.org/c/s/www.brookings.edu/research/a-guide-to-healthy-skepticism-of-artificial-intelligence-and-coronavirus/amp/.

[5] Yuan, Ye, Xu, Hui. Dr., Shusheng Li, Dr. "A machine learning-based model for survival prediction in patients with severe COVID-19 infection." MedRxiv, Mar. 17, 2020, available at https://www.medrxiv.org/content/10.1101/2020.02.27.20028027v3.full.pdf.

[6] Lopez, C. Todd. "DOD Adopts 5 Principles of Artificial Intelligence Ethics." DOD News, U.S. Dept. of Defense, Feb. 25, 2020, available at https://www.defense.gov/Explore/News/Article/Article/2094085/dod-adopts-5-principles-of-artificial-intelligence-ethics/.

[7] Algorithmic Accountability Act of 2019, H.R. 2231, 116 Cong. (1st Sess. 2019), available

at https://www.wyden.senate.gov/imo/media/doc/Algorithmic%20Accountability%20Act%20of%202019%20Bill%20Text.pdf.

[8] Commercial Facial Recognition Privacy Act of 2019, S. 847, 116 Cong. (1st Sess. 2019), available at https://www.congress.gov/116/bills/s847/BILLS-116s847is.pdf.

[9] Henry, Linda. "NYC's Task Force to Tackle Algorithmic Bias Issues Final Report." JD Supra, Patrick Law Group, LLC, Jan. 31, 2020, available at https://www.jdsupra.com/legalnews/nyc-s-task-force-to-tackle-algorithmic-93703/.

[10] Tiku, Nitasha. "Microsoft Wants to Stop AI's Race to the bottom." Wired.com, Dec. 6, 2018, available at https://www.wired.com/story/microsoft-wants-stop-ai-facial-recognition-bottom/.

[11] European Commission, White Paper Report "On Artificial Intelligence – A European approach to excellence and trust." Brussels, Feb. 19, 2020, available at https://ec.europa.eu/info/sites/info/files/commission-white-paper-artificial-intelligence-feb2020_en.pdf.

[12] IBM Research Trusted AI, "AI Fairness 360 Open Source Toolkit." Retrieved ___. __, 2020, available at http://aif360.mybluemix.net/#.

[13] Ibid., at https://pypi.org/project/aif360/.

[14] LaBrie, Ryan C., Steinke, Gerhard H. "Towards a Framework for Ethical Audits of AI Algorithms" Emergent Research Forum, Twenty-fifth Americas Conference on Information Systems, 2019, available at https://aisel.aisnet.org/cgi/viewcontent.cgi?article=1398&context=amcis2019.

[15] Wilson, James H., Daugherty, Paul R., Bianzini-Morini, Nicola. "The Jobs That Artificial Intelligence Will Create." MIT Sloan Management Review, Mar. 23, 2017, available at https://sloanreview.mit.edu/article/will-ai-create-as-many-jobs-as-it-eliminates/.

[16] Rosén, Josefin. "What every business manager should know about algorithm audits." SAS, Hidden Insights (Blog), Oct. 16, 2017, available at https://blogs.sas.com/content/hiddeninsights/2017/10/16/algorithm-audits/.

[17] Bostrom, Nick, Yudkowsky, Eliezer. "The Ethics of Artificial Intelligence." The Cambridge Handbook of Artificial Intelligence, New York: Cambridge University Press. 2014, available at https://intelligence.org/files/EthicsofAI.pdf.

[18] Davenport, Thomas H. "What Does an AI Ethicist Do?" MIT Sloan Management Review, Jun. 24, 2019, available at https://sloanreview.mit.edu/article/what-does-an-ai-ethicist-do/?gclid=CjwKCAjwqJ_1BRBZEiwAv73uwCgKxdJ0lm9o8bZdvt8_FNMQeLUerfeWyKirn3pQrHrhtV9NrbvrVhoCA_kQAvD_BwE.

[19] Lopez, Todd (2020).