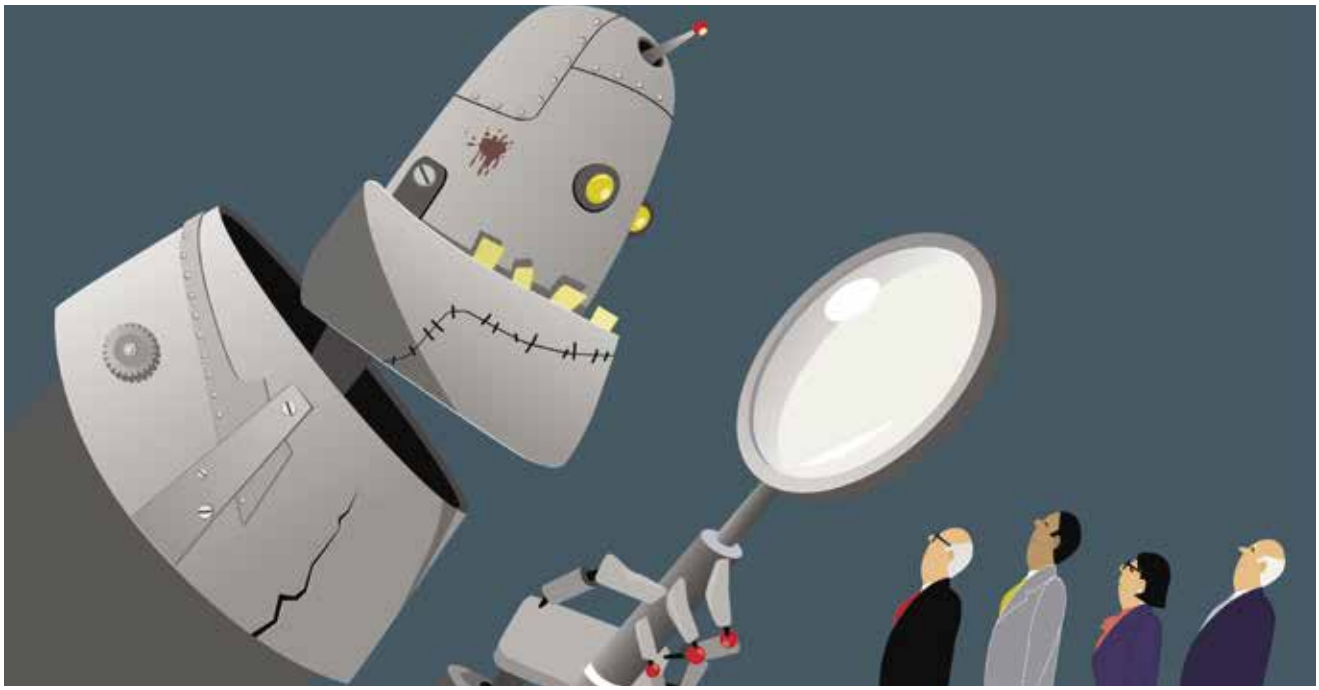
Artificial Intelligence and Bias

By Judge Herbert B. Dixon Jr.



I am an unabashed supporter of artificial intelligence (AI); however, all users should be aware of the potential limitations of AI applications.

I have written three technology columns discussing AI limitations in the past year. In the article "My 'Hallucinating' Experience with ChatGPT,"[1] I discussed the propensity of some artificial intelligence models to hallucinate their response, that is, make up facts, events, and publications to support an answer provided by the AI. In the article "Artificial Intelligence: 'What Hath God Wrought,'"[2] I discussed concerns that AI platforms' chatbots could become more intelligent than humans and be exploited by bad actors, the ability of AI to spread misinformation, and the fears that AI users may be unaware that private personal and business information they submit for analysis by AI platforms may end up in the public domain. In my most recent AI article, "Artificial Intelligence versus Copyright Protections and Data Privacy,"[3] I discussed concerns by copyright holders that some AI applications were being trained on copyrighted materials, which would then be improperly used without the copyright holder's permission, in responses provided to the AI users. I also discussed concerns about private personal and business information stored within AI platforms without adequate security. This article aims to heighten readers' awareness of another AI limitation, "bias."

As used in this article, bias is similar to the Wikipedia definition, namely, a disproportionate weighting of factors in an unfair way, using underrepresented samples of a population, or an estimation process that does not give accurate results on average.[4]

What do you think the response would be if you asked 100 random people the following question? Which do you think is most capable of analyzing data and providing an accurate and impartial answer, a human or a computer? I suspect an overwhelming number of responding people would choose the computer. Why? Because most people think of computers as unemotional neutral devices that respond with answers based on facts and mathematical computations. The proposed question is probably wrong because humans create computers and design and train the systems to make them work. As

Judge Herbert B. Dixon Jr. is a senior judge with the Superior Court of the District of Columbia. He is chair of the *ABA Journal* Board of Editors, a former chair of both the National Conference of State Trial Judges and the ABA Standing Committee on the American Judicial System, and a former member of the Techshow Planning Board. You can reach him at Jhbdixon@gmail.com. Follow Judge Dixon on X (formerly known as Twitter) @Jhbdixon.

these computer systems are created, the argument goes, they reflect the biases of their human creators.[5] Wow!

The idea of AI bias is not a wild or fringe concept. The subject has been discussed extensively in scholarly writing, news articles, and practical "how-to" pieces, and the existence of AI bias has been documented within numerous AI applications.

First, I must caution readers. This article emphasizes early and sometimes worse-case scenarios of AI bias. However, every year, there has been improvement in limiting the bias effect in AI applications. As I discuss instances of AI bias below, understand that today's concerns may not be the same as tomorrow's. While improvements in AI results are occurring continuously, AI creators and users must be vigilant to avoid the law of unintended consequences, such as the solution to one problem causing a new problem, or when one problem is solved, another problem that previously went unnoticed becomes apparent.

Well-documented instances of AI bias occurred within Amazon's corporate structure, which has been remarkably successful in using AI to analyze its customers' purchases, make predictions regarding their future needs, and create more efficiencies with their prominent use of robots in the Amazon distribution centers. Amazon scrapped two separate AI personnel tools. One tool was used to review résumés in an effort to find the best job candidates. Eventually, Amazon abandoned the tool because it was biased against women. The AI tool was trained on 10-year-old information about Amazon's previous hires in which women were underrepresented.[6] Consequently, the AI application favored men over women.

Amazon made another effort when it attempted to develop an AI tool to autonomously search the internet for candidates deemed worthy of recruitment. The team created 500 computer models to recognize 50,000 terms in past candidates' résumés. Eventually, this project was abandoned. An example of the bias discovered in the computer models was that it assigned a higher value to terms such as "executed" and "captured," which were commonly found in the résumés submitted by male engineers. In addition to instances of gender bias, and apparently based on the data used to train the computer models, the AI application often recommended unqualified candidates for jobs and, in some instances, seemed to recommend candidates at random.[7]

Facial recognition technology is an example of AI technologies that started with severe criticisms because of biased results, which have improved substantially since its early days, according to research by the National Institute of Standards and Technology (NIST).[8] One researcher uncovered significant gender and racial bias in early facial recognition AI systems in that they performed substantially better on male faces than female faces. In addition, the systems had error rates of about 1 percent for lighter-skinned men and 35 percent for darker-skinned women. Also of note, several early systems failed to correctly classify the faces of Oprah Winfrey, Michelle Obama, and Serena Williams, for example, identifying them as a gentleman, young boy, or young man, and misinterpreting their hair as a cap or a headpiece.[9]

Although facial recognition is now a feature of most smartphones and the software's accuracy today is much improved, the use of facial recognition technology just a few years ago was the subject of significant controversy due to the misidentification of Black people compared to whites. In 2019, San Francisco banned law enforcement's use of facial recognition technology. Within two years after the San Francisco ban, at least 16 municipalities enacted similar local bans, primarily because of perceived AI bias. Indeed, California enacted a three-year statewide ban on the use of the technology starting in January 2020.[10] This year, in January 2024, legislative bills were filed in the New York Assembly and Senate banning law enforcement's use of facial recognition and other biometric surveillance technology.[11] However, due to improvements in the technology, the opposition now centers more on individual privacy rather than AI bias.

AI risk assessment tools for defendants facing criminal charges and predictive AI applications for police patrol practices also have been criticized. Algorithms used for risk assessments to predict the chances that the individual would commit another crime were later assessed to have biases against Black people. Unfortunately for the subjects of those risk assessments, the scores were used to determine bail, sentencing, and parole.[12]

Predictive AI applications are used to identify specific areas as hot spot locations for officers to expect trouble when on patrol, and, based on the predictions, police will allocate resources to meet the anticipated need. The early predictive AI applications were trained on historical data, which, not surprisingly, attributed priority to the areas where more arrests historically occurred. Unfortunately, practice indicates that additional police resources in those areas increased the likelihood that police would stop or arrest people in the same locations, thus reinforcing the historical pattern and whatever biases were baked into the training data.[13]

An often-cited 2016 investigative article in *ProPublica* reviewed several instances of risk assessment scores that were questionable. One of the cited risk score discrepancies on which the article based its conclusions of AI bias concerned the cases of Brisha Borden and Vernon Prater, who were arrested on separate occasions in Florida. Borden, an 18-year-old Black female, was arrested when she and a friend grabbed an unlocked bicycle and scooter in their neighborhood and started to ride away. Before they got away, the two were arrested and charged with burglary and petty theft of items valued at $80. Borden had four juvenile misdemeanor offenses in her past. She received a high-risk assessment score of 8 out of 10. Prater, a 41-year-old white male, was picked up for shoplifting $86.35 worth of tools from a nearby Home Depot store. Prather's previous record included an attempted armed robbery offense and two armed robbery offenses for which he

served five years in prison. Prather received a low-risk assessment score of 3 out of 10. Everyone agreed the AI risk assessment got it wrong. Borden had no subsequent offense, but Prater had a subsequent offense of grand theft.[14]

Final Comments

Again, as expressed above, this article aims to raise the readers' consciousness and awareness to be ever diligent about unquestioningly accepting AI results. To the extent that AI bias exists, Sam Altman, the founder and CEO of OpenAI, the creator of ChatGPT, has offered his opinion that AI systems will eventually fix themselves. He bases his conclusion on a technology called RLHF (reinforcement learning from human feedback). Altman accepts that early AI systems may have reinforced their creators' biases.[15]

Unfortunately, however, even if we reach the point of AI systems correcting their built-in biases, there is another type of bias that could defeat that outcome even if the AI application provides an accurate prediction, namely, human review bias. That includes a situation where the AI makes a correct prediction, but the human reviewer negates the result with their own bias. For example, the human reviewer concludes the AI prediction is wrong because the human reviewer knows the people in the designated neighborhood and is confident they would never act the way the AI system predicts. In such a situation, all the work done to produce an AI that correctly predicts the results would be undone by human review bias, the same issue that started us looking to computers to produce a fairer result. ■

Endnotes

1. Herbert B. Dixon Jr., My "*Hallucinating*" *Experience with ChapGPT*, 62 Judges' J., no. 2, Spring 2023, at 37, https://bit.ly/3O8aPiB.

2. Herbert B. Dixon Jr., *Artificial Intelligence: What Hath God Wrought,"* 62 Judges' J., no. 3, Summer 2023, at 37, https://bit.ly/48E2r2g.

3. Herbert B. Dixon Jr., *Artificial Intelligence versus Copyright Protections and Data Privacy*, 62 Judges' J. no. 4, Fall 2023, at 37, https://bit.ly/48E2zyM.

4. *Bias*, Wikipedia (Jan. 8, 2024), https://en.wikipedia.org/wiki/Bias.

5. A.W. Ohlheiser, *AI Automated Discrimination. Here's How to Spot It. The Next Generation of AI Comes with a Familiar Bias Problem*, Vox (June 14, 2023), https://bit.ly/41PVqZQ.

6. Jeffrey Dastin, *Amazon Scraps Secret AI Recruiting Tool That Showed Bias Against Women*, Reuters (Oct. 10, 2018), https://reut.rs/2u0RJo5.

7. *Id.*

8. Titus Wu, *California at Crossroads over Policing and Facial Recognition*, Bloomberg Law (Mar. 29, 2023), https://bit.ly/3Hdr9dR.

9. Joy Buolamwini, *Artificial Intelligence Has a Problem with Gender and Racial Bias. Here's How to Solve It*, Time (Feb. 7, 2019), https://bit.ly/3HgbhY9.

10. Nathan Sheard & Adam Schwartz, *The Movement to Ban Government Use of Face Recognition*, Elec. Frontier Found. (May 5, 2022), https://bit.ly/3RTWqHz.

11. Mike Maharrey, *New York Bills Would Ban Law Enforcement Use of Facial Recognition Surveillance*, Tenth Amend. Ctr. (Jan. 8, 2024), https://bit.ly/48tcDut.

12. Ohlheiser, *supra* note 5.

13. *Id., supra note 5.*

14. Julia Angwin et al., *Machine Bias. There's Software Used Across the Country to Predict Future Criminals. And It's Biased Against Blacks*, ProPublica (May 23, 2016), https://bit.ly/41RmCHr.

15. David I. Adeleke, *AI Models Can Be a Force to Reduce Bias, Not Reinforce It, Sam Altman Says*, Rest of World (May 23, 2023), https://bit.ly/4aR47af.