

# Bias in Artificial Intelligence

Gregory S. Nelson

Imagine an algorithm that selects nursing candidates for a multi-specialty practice—but it only selects white females. Consider a revolutionary test for skin cancer that does not work on African Americans. What about a model that directs poorer patients to a skilled nursing facility rather than their home as it does for wealthier patients? These are ways in which ungoverned artificial intelligence (AI) might perpetuate bias.

With the current hyperbole around AI approaching an all-time high, it takes little imagination to see how the algorithms applied in other industries can be used in health care. Google algorithms for automated image classification can be modified to read CT scans of a patient with cancer [1] or predict treatable blinding retinal diseases [2]. The AI methods used to predict the risk of loan default can be tweaked to predict the risk of sepsis [3, 4] or pneumonia [5].

In health care, clinical decision support has long been integrated into our electronic health records to guide safe medication use, use of clinical best practices, and prioritization of high-risk patients. We could, of course, choose to ignore any algorithm's suggestions much like we might snub Amazon's book recommendations. However, when AI systems go beyond recommendations and act autonomously, we must pause and consider the implications. At best, we streamline processes, reduce variation in care, and remove human biases from decision-making [6-8]. At worst, we erode trust, perpetuate gender, ethnicity, and income disparities, and distance ourselves from patient care decisions.

There is ample evidence of bias in AI [9]. Also known as algorithmic bias, it is what we experience when a machine-learning model produces a systematically wrong result. Just as this article is a reflection of the bias of its author, algorithms have authors and are assembled according to instructions made by people. Bias is a reflection of the data algorithm authors choose to use, as well as their data blending methods, model construction practices, and how results are applied and interpreted. That is to say, these processes are driven by human judgments.

Health care is one of the most challenging industries when it comes to data, primarily due to the fact that the industry's operational systems were not designed for modern analytics and are often not fully integrated with internal or external data systems. We are still learning about the full spectrum of factors that determine health outcomes [10-12]. Sadly, most health care organizations still grapple with issues like data quality, data governance, and effective use of Health IT to improve outcomes. That is to say, we may use the data that we have as opposed to the data that is "right".

We are plagued with data that cannot be integrated across people, time, or place; information collected for billing purposes which does not fully reflect the underlying diagnosis and treatment [13, 14]; and data collection practices that are highly biased toward those who can afford health care services. We often see this manifested in patients' access to care where the data in the EHR can be shallower for some segments of the population [15] or in curated health care data that is resold by brokers where

bias exists toward those who can afford devices, applications, and technology [16].

As we evaluate sources of bias in our models, it is essential that we establish principles to guide our work. We must adopt four primary tenets: transparency, trust, fairness, and privacy.

Transparency stresses the responsibility of AI authors to explain not only what went into an algorithm and its results, but what decisions they made and why. The goal is to understand the process by which an algorithmic system makes decisions, and we must ensure the model can be explained. Often called the “black-box problem,” this challenge often poses issues for physicians who seek insight into what the AI is doing.

Trust begins with transparency, verification, and accountability. As Dr. Wyatt Decker, the Mayo Clinic’s chief medical information officer, points out: “... clinician involvement is important no matter how smart the machines get. There is a strong need for the engagement of medical experts to validate and oversee AI algorithms in healthcare” [17].

“Fairness” is a social construct, and in the context of bias in AI we are referring to being responsible for social mores. Algorithms are discriminatory in that they seek tiny patterns of influence in the data. Anthropomorphically speaking, we want a model that is socially responsible—one that does not discriminate against people based on traits that we would generally consider protected (eg, age, gender, sexual orientation, race, or ethnicity).

Privacy reflects on the nature of our relationship with our patients. While there are certainly cases of people using geospatially derived variables, purchase history, and social media data to augment the medical records, we must ensure the protection of individual privacy at all times.

There is a growing body of work in the legal [18], regulatory [19], and ethical oversight of AI models [20]. These sources ask that we look beyond the technical processes for data selection, model building, and validation and adopt formal AI governance strategies. In this context, AI governance is in the process of assigning and assuring organizational accountability, decision rights, risks, policies, and investment decisions for applying artificial intelligence. Newly proposed federal legislation, The Algorithmic Accountability Act of 2019, would require businesses to conduct an impact assessment that covers the risk associated with algorithms’ accuracy, fairness, bias, discrimination, privacy, and security [21]. Surely health care will fall under the umbrella of such legislation. Research and consulting firm Gartner, Inc. predicts that by 2022 the first US medical malpractice case involving a medical decision made by an advanced AI algorithm will have been heard [22]. It will not be because an algorithm produced an incorrect diagnosis. “It will be due to the failure to use an algorithm that was proven to be more accurate and reliable than the human alone [22].”

It is our professional and moral obligation to do what we can to ensure that AI is safe for our patients and care teams. Given the tenets of trust, fairness, transparency, and privacy, we should focus on solutions that may help care teams automate the activities that take them away from their patients, not on replacing them. Continued advancement of AI in health care will require stakeholder education and the management of expectations so that we can eradicate unintentional bias and engender trust in transparent, clinically validated models. After all, as Manu Tandon, chief information officer of Beth Israel Deaconess Medical Center in Boston, suggests, “We are not looking for robots to do work for us, we are looking to make

## better decisions by benefiting from machine learning and AI" [17]. NCMJ

**Gregory S. Nelson, MMCi, CPHIMS** vice president of analytics and strategy, Vidant Health, Greenville, North Carolina; adjunct faculty, Duke University Fuqua School of Business, Duke University, Durham, North Carolina.

### Acknowledgments

Potential conflicts of interest. G.S.N. has no relevant conflicts of interest.

### References

1. Thompson RF, Valdes G, Fuller CD, et al. Artificial intelligence in radiation oncology: a specialty-wide disruptive transformation? *Radiother Oncol.* 2018;129(3):421-426.
2. Kermany DS, Goldbaum M, Cai W, et al. Identifying medical diagnoses and treatable diseases by image-based deep learning. *Cell.* 2018;172(5):1122-1131.
3. Calvert JS, Price DA, Chettipally UK, et al. A computational approach to early sepsis detection. *Comput Biol Med.* 2016;74:69-73.
4. McCoy A, Das R. Reducing patient mortality, length of stay and readmissions through machine learning-based sepsis prediction in the emergency department, intensive care unit and hospital floor units. *BMJ Open Qual.* 2017;6(2):e000158.
5. Caruana R, Lou Y, Gehrke J, Kock P, Sturm M, Elhadad N. Intelligible models for healthcare: predicting pneumonia risk and hospital 30-day readmission. *KDD '15 Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining.* Sydney, NSW, Australia: ACM; 2015. [https://www.microsoft.com/en-us/research/wp-content/uploads/2017/06/KDD2015FinalDraftIntelligibleModels4HealthCare\\_igt143e-caruanaA.pdf](https://www.microsoft.com/en-us/research/wp-content/uploads/2017/06/KDD2015FinalDraftIntelligibleModels4HealthCare_igt143e-caruanaA.pdf). Accessed May 7, 2019.
6. Tversky A, Kahneman D. Judgment under uncertainty: heuristics and biases. *Science.* 1974;185(4157):1124-1131.
7. Kahneman D, Slovic SP, Tversky A. *Judgment Under Uncertainty: Heuristics and Biases.* Cambridge, NY: Cambridge University Press; 1982.
8. Kahneman D, Tversky A. *Choices, Values, and Frames.* New York, NY: Cambridge University Press; 2000.
9. Yu KH, Kohane IS. Framing the challenges of artificial intelligence in medicine. *BMJ Qual Saf.* 2019;28(3):238-241.
10. Artiga S, Hinton E. *Beyond Health Care: The Role of Social Determinants in Promoting Health and Health Equity.* Menlo Park, CA: Henry J Kaiser Family Foundation; 2018.
11. Braveman P, Gottlieb L. The social determinants of health: it's time to consider the causes of the causes. *Public Health Rep.* 2014;129(suppl 2):19-31.
12. Hernandez LM, Blazer DG, Institute of Medicine (US) Committee on Assessing Interactions Among Social, Behavioral, and Genetic Factors in Health. *Genes, behavior, and the social environment: moving beyond the nature/nurture debate.* Washington, DC: National Academies Press; 2006.
13. Pine K, Mazmanian M. Institutional Logics of the EMR and the problem of 'perfect' but inaccurate accounts. 17th ACM Conference on Computer Supported Cooperative Work and Social Computing. Baltimore, MD: CSCW; 2014. <https://asu.pure.elsevier.com/en/publications/institutional-logics-of-the-emr-and-the-problem-of-perfect-but-in>. Accessed May 7, 2019.
14. Alert coders about new guidance for coding body mass index and pressure ulcers. *Medical Coding World website.* <https://medicalcodingpro.wordpress.com/2009/04/22/alert-coders-about-new-guidance-for-coding-body-mass-index-and-pressure-ulcers/>. Published April 22, 2009. Accessed May 6, 2019.
15. Han X, Call KT, Pintor JK, Alarcon-Espinoza G, Simon AB. Reports of insurance-based discrimination in health care and its association with access to care. *Am J Public Health.* 2015;105(suppl 3):517-525.
16. Jeong S. Insurers Want to Know How Many Steps You Took Today. *NYTimes.com.* <https://www.nytimes.com/2019/04/10/opinion/insurance-ai.html>. Published April 10, 2019. Accessed May 16, 2019.
17. Arndt RZ. The slow upgrade to artificial intelligence. *ModernHealthcare.com.* <https://www.modernhealthcare.com/indepth/artificial-intelligence-in-healthcare-makes-slow-impact/>. Accessed May 6, 2019.
18. Coglianese C, Lehr D. Transparency and algorithmic governance. *Public Law Research Paper, No. 18-38.* Administrative Law Review, U of Penn Law School. 2018;71:1.
19. U.S. Food & Drug Administration. Proposed Regulatory Framework for Modifications to Artificial Intelligence/Machine Learning (AI/ML)-Based Software as a Medical Device (SaMD): Discussion Paper and Request for Feedback. Washington, DC: FDA; 2019. <https://www.fda.gov/media/122535/download>. Accessed May 7, 2019.
20. Floridi L, Cowls J, Beltrametti M, et al. AI4People—an ethical framework for a good AI society: opportunities, risks, principles, and recommendations. *Minds and Machines.* 2018;28(4):689-707.
21. Algorithmic Accountability Act, 116th Cong, 1st Sess (2019).
22. Pessin G, Lovelock JD, Runyon B, et al. Predicts 2019: Healthcare Providers Must Embrace Digital Transformation. *Gartner, Inc. website.* <https://www.gartner.com/en/documents/3895268>. Published December 10, 2018. Accessed May 7, 2019.

Electronically published July 8, 2019.

Address correspondence to Greg Nelson, 2100 Stantonsburg Road, Greenville, NC 27835-6028 (greg.nelson@vidantthehealth.com).

**N C Med J.** 2019;80(4):220-222. ©2019 by the North Carolina Institute of Medicine and The Duke Endowment. All rights reserved. 0029-2559/2019/80407