ImproveDx Newsletter
May 2021 - Volume 8 | Issue 3

Bias in Artificial Intelligence

> *…inequity and bias are not to be found in a single place, like a bug that can be located and fixed. These issues are systemic.* [1(p9)]

Recent news stories deliver apparently contradictory messages about artificial intelligence (AI): future prospects leading to lucrative business deals on the one hand and disappointing performance prompting protests and lawsuits on the other.[2-4] Expectations remain high that AI will continue to transform many aspects of daily life, including health care — thus the business deals, such as Microsoft's commitment to acquire Nuance Communications for $19.7 billion.[5] At the same time, AI algorithms have produced results biased against women and people of color, prompting disillusionment and reassessment.

Racism and Gender Inequities

Troubling examples demonstrate how AI can reflect and even exacerbate existing racism and gender inequities. In 2015, Black software developer Jacky Alciné was shocked to find that Google Photos had automatically sorted selfies he'd taken with a friend, who is also Black, into a default folder labeled "gorillas."[6,7] And in 2017, Joy Buolamwini, a Ghanaian-American computer scientist at the MIT Media Lab could only get a facial recognition program to see her by putting on a white mask.[8,9] Other examples abound.

The source of algorithmic bias is more implicit than explicit. It is inherent in the environment within which the technology was developed. Both of the above examples involve facial recognition, but the underlying problem of faulty, biased and offensive results occurs across many algorithms and applications of AI.

Machine learning, the process through which algorithms learn to identify patterns in data, relies on vast digital archives of tagged images, records, and other data for continual training and refinement. Most of the online information currently available in the quantities AI craves is biased toward white males, with people of other races, sex and gender identification greatly underrepresented.

The bias toward white males is persistent and systemic, inherent in the training data and carried forward in algorithms that can accentuate and intensify the bias. Buolamwini warns, "Algorithms, like viruses, can spread bias on a massive scale, at a rapid pace."[10(np)]

Algorithms exert this power because they are embedded in systems everywhere, often generating influential results without revealing how they arrive at their conclusions.[9] In his book on artificial intelligence in medicine,[11] Eric Topol, MD, casts modern day algorithms as agents, operating in the world according to human-written instructions, with growing independence:

> *Algorithms have thus become agents…. Algorithms now do things.…They drive cars. They manufacture goods. They decide whether a client is creditworthy.*[12]

They also help diagnose diseases. Some industries look forward to the day when algorithms act autonomously — in self-driving automobiles, for example. Despite futuristic projections about physicians being replaced by algorithms, in healthcare AI remains an assistive technology, with growing influence but no real prospect for independent decision-making in the foreseeable future.[11,13]

Another example of bias in healthcare demonstrates how past inequities can play forward in these models. While analyzing hospital data to measure the impact of a managed care program, researchers were surprised to find that among patients with comorbidities, Black patients were on average assigned lower risk scores than white patients who appeared to have similar medical conditions.[14] And those lower scores meant the Black patients would receive less personalized care.

Digging deeper, researchers discovered that the algorithm assigned risk scores based on the patient's annual cost of care. That is to say, the developers used cost of care as a proxy for complexity of medical condition. But Black patients often receive less care for a variety of reasons related to systemic racism: poor access to care, inadequate health insurance, lack of respect, distrust of the health care system, and other barriers to receiving care. It turned out that Black patients who had similar health care costs as white patients had more comorbid conditions and were sicker.

Hospitals and insurers in the U.S. use the algorithm involved in this study to manage care for approximately 200 million people annually. After this story became known through an article in *Science,* U.S. Senators Cory Booker (D-NJ) and Ron Wydon (D-OR) urged the Centers for Medicare & Medicaid Services, the Federal Trade Commission and other leading health care payers and regulators to address the problem of bias in health-care AI.[15] The algorithm's developer is now working with the research team to correct the problem.

Interviewed when the study was published in 2019, lead author Ziad Obermeyer, MD, reflected on the challenge of undoing systemic bias in AI:

> *Those solutions are easy in a software-engineering sense: you just rerun the algorithm with another variable.… But the hard part is: what is that other variable? How do you work around the bias and injustice that is inherent in that society?*[16(608)]

If the bias embedded in these systems and the resulting harm were simply caused by glitches or homogeneity in the training data or through a lack of awareness among computer scientists, it

would be easier to fix. Fundamentally, the problem reflects long-standing prejudice and discrepancies of power throughout society, including the corporations that develop and benefit from AI services and products, the educational institutions that train scientists, the research community, public and private investors, and so on.[1] As with other issues that stem from systemic prejudice related to race, ethnicity, sex, and gender, awareness is the first step in a long journey toward inclusion, equity, and fairness.

AI Bias and Diagnosis

In medicine, specialties that rely on processing visual information and pattern recognition skills — dermatology, radiology, ophthalmology, and pathology — are among early adopters of AI systems designed to assist with diagnosis.

Because skin color affects the presentation of conditions and diseases, dermatology's experience with AI includes working with racial differences and the potential for bias.[17,18]

VisualDx, provider of diagnostic clinical decision support to advance pattern recognition in medicine and dermatology, has been aggregating images of disease presentations in people of color for over 20 years. The images are classified by diagnosis and by Fitzpatrick Skin Type, a standard phototype categorization used often to define pigmentation. VisualDx's CEO Art Papier, MD, explains, "Erythema (skin redness) and purpura (a sign of blood leakage out of blood vessels), for example, are physical exam clues relatively easy to see on white skin, but they offer a different challenge on darker skin tones." He adds, "Machine learning is completely dependent on training algorithms on excellent data. In dermatology, as in all the visual specialties, the effectiveness of machine learning requires highly reliable 'ground truth' for training" (written communication, April 2021).

In radiology, recent research shows the effect that biased training images can have. Researchers in Argentina evaluated the impact of gender imbalance in imaging archives used to train AI-based systems for computer-aided diagnosis (CAD).[19] They found reduced diagnostic performance for female patients when the CAD system had been trained on images from male patients. Running a number of different scenarios for gender balance—0% women/100% men, 25%/75%, 50%/50%—they consistently found that imbalance in the training data negatively impacted the accuracy of results. The best performance for both male and female patients came from a system trained with data balanced 50/50 for gender.

In 2019, a consortium of organizations, including the American College of Radiology, RSNA, American Association of Physicists in Medicine, and imaging societies in Canada and Europe, developed a joint statement to address bias and other ethical issues in the growing use of AI in radiology. The statement proposes a set of 8 questions that those responsible for AI systems should be able to answer, including:

- What kinds of bias may exist in the data used to train and test algorithms?
- What have we done to evaluate how data are biased, and how it may affect our model?

- What are the possible risks that might arise from biases in the data? [20(p438)]

These and the other questions proposed are crucial for ethical use of these systems, but the ability to supply answers and guidance for future practice is still a work in progress.

Transparency

AI supplies systems that are sophisticated, complex, and often inscrutable even to the computer scientists who develop them. Describing Jacky Alciné's experience with Google Photos, which Google was able to fix only by removing "gorilla" as a category, computer scientist Erik Larson comments:

> *You have these terrible, insensitive results from the system, and the system just doesn't know what it's doing. We don't know what it was focusing on or how it made that decision. These systems are notoriously opaque…you can't deconstruct them after the fact.*[21(31:31 mins)]

Beyond the outrage prompted by results that are obviously insulting or subtly discriminatory, AI results that are erroneous and unfair undermine the public's trust. Given general agreement that AI will ultimately improve the quality and safety of health care,[22] maintaining the public's trust is another reason to address problems with bias.

There are many efforts underway to deal with the problem. Some are calling for AI ethics to be included in medical school education.[23] Others suggest that trying to fix bias within AI technology is a "seductive diversion"[24] from dealing with questions of racism, corporate power, and data ownership. And there is a movement to create "explainable" artificial intelligence to design transparency and accountability into these systems.

In recent remarks, Micki Tripathi, PhD, President Biden's National Coordinator for Health IT, included AI bias among the agency's top priorities. Acknowledging a history of policies that have been advanced without understanding the possible effects across all populations, he expressed a desire to take health equity into account up front. Forecasting that this will continue to be an issue of growing concern, Tripathi said,

> *I think everyone's familiar with the issues with algorithmic bias. And that is a bigger and bigger issue the more that algorithms are embedded in…all the technologies that we use.*[25]

The pandemic has focused attention on systemic racial injustice and health care disparities. And it has accelerated adoption of remote technologies often enhanced with artificial intelligence. Although the challenges are formidable, there is a growing movement to address the problem of bias in AI and be able to harness its potential to improve diagnosis and other aspects of health care without causing further societal division and harm.

Thank you to our reviewers: Jen J. Gong, PhD; Art Papier, MD; Lorri Zipperer, MA

References

1. West SM, Whittaker M, Crawford K. *Discriminating Systems: Gender, Rage, and Power in AI.* AI Now Institute; 2019. https://ainowinstitute.org/discriminatingsystems.pdf
2. Roosli E, Rice B, Hernandez-Boussard T. Bias at warp speed: how AI may contribute to the disparities gap in the time of COVID-19. *J Am Med Inform Assoc.* 2021;28(1):190-192. doi:10.1093/jamia/ocaa210
3. Hao K. We read the paper that forced Timnit Gebru out of Google. Here's what it says. *MIT Technology Review.* December 4, 2020. https://www.technologyreview.com/2020/12/04/1013294/google-ai-ethics-research-paper-forced-out-timnit-gebru/
4. Ryan-Mosley T. The new lawsuit that shows facial recognition is officially a civil rights issue. *MIT Technology Review.* April 4, 2021. https://www.technologyreview.com/2021/04/14/1022676/robert-williams-facial-recognition-lawsuit-aclu-detroit-police/.
5. Bray H. In $19.7 billion deal for Nuance, Microsoft sees big opportunity in health care and artificial intelligence. *Boston Globe.* April 12, 2021. https://www.bostonglobe.com/2021/04/12/business/microsoft-buy-burlington-technology-firm-nuance-communications-197-billion/.
6. Simonite T. When it comes to gorillas, google photos remains blind. *Wired.* January 11, 2018. https://www.wired.com/story/when-it-comes-to-gorillas-google-photos-remains-blind/
7. Barr A. Google mistakenly tags Black people as 'gorillas,' showing limits of algorithms. *Wall Street Journal.* July 1, 2015. https://www.wsj.com/articles/BL-DGB-42522
8. Lee J. When bias is coded into our technology. *Code Switch.* NPR. February 8, 2020. https://www.npr.org/sections/codeswitch/2020/02/08/770174171/when-bias-is-coded-into-our-technology
9. Metz C. Who is making sure the A.I. machines aren't racist? *New York Times.* March 15, 2021. https://www.nytimes.com/2021/03/15/technology/artificial-intelligence-google-bias.html
10. Buolamwini J. How I'm fighting bias in algorithms. TED. March 29, 2017. https://www.youtube.com/watch?v=UG_X_7g63rY
11. Topol E. *Deep Medicine: How Artificial Intelligence Can Make Healthcare Human Again.* Basic Books; 2019.
12. Mazzotti M. Algorithmic life. *Los Angeles Review of Books.* January 22, 2017. https://lareviewofbooks.org/article/algorithmic-life/
13. Mukherjee S. A.I. versus M.D. *New Yorker.* March 27, 2017. https://www.newyorker.com/magazine/2017/04/03/ai-versus-md

14. Obermeyer Z, Powers B, Vogeli C, Mullainathan S. Dissecting racial bias in an algorithm used to manage the health of populations. *Science.* 2019;366(6464):447-453. doi:10.1126/science.aax2342

15. Eddy N. Senators urge health orgs to combat racial bias in AI algorithms. *Healthcare IT News.* December 6, 2019. https://www.healthcareitnews.com/news/senators-urge-health-orgs-combat-racial-bias-ai-algorithms

16. Ledford H. Millions of black people affected by racial bias in health-care algorithms. *Nature.* 2019;574(7780):608-609. doi:10.1038/d41586-019-03228-6

17. Alvarado SM, Feng H. Representation of dark skin images of common dermatologic conditions in educational resources: A cross-sectional analysis. *J Am Acad Dermatol.* 2021;84(5):1427-1431. doi:10.1016/j.jaad.2020.06.041

18. Ebede T, Papier A. Disparities in dermatology educational resources. *J Am Acad Dermatol.* 2006;55(4):687-690. doi:10.1016/j.jaad.2005.10.068

19. Larrazabal AJ, Nieto N, Peterson V, Milone DH, Ferrante E. Gender imbalance in medical imaging datasets produces biased classifiers for computer-aided diagnosis. *Proc Natl Acad Sci USA.* 2020;117(23):12592-12594. doi:10.1073/pnas.1919012117

20. Geis JR, Brady AP, Wu CC, et al. Ethics of artificial intelligence in radiology: summary of the Joint European and North American Multisociety Statement. *Radiology.* 2019;293(2):436-440. doi:10.1148/radiol.2019191586

21. Interview with Erik Larson. The myth of artificial intelligence. The Lawfare Podcast. March 31, 2021. https://www.lawfareblog.com/lawfare-podcast-myth-artificial-intelligence

22. Bates DW, Levine D, Syrowatka A, et al. The potential of artificial intelligence to improve patient safety: a scoping review. *NPJ Digit Med.* 2021;4(1):54. doi:10.1038/s41746-021-00423-6

23. Katznelson G, Gerke S. The need for health AI ethics in medical school education [published online ahead of print March 3, 2021.] *Adv Health Sci Educ Theory Pract.* 2021. doi:10.1007/s10459-021-10040-3

24. Powles J, Nissenbaum H. The seductive diversion of 'solving' bias in artifical intelligence. OneZero. December 7, 2018. https://onezero.medium.com/the-seductive-diversion-of-solving-bias-in-artificial-intelligence-890df5e5ef53.

25. Miliard M. ONC chief Micky Tripathi talks public health data systems and 'health equity by design.' *Healthcare IT News.* March 12, 2021. https://www.healthcareitnews.com/news/onc-chief-micky-tripathi-talks-public-health-data-systems-and-health-equity-design

Working Together to Improve Diagnosis: Virtual Conference

Have you heard? The Society to Improve Diagnosis in Medicine (SIDM) is partnering with Constellation® and Iowa Healthcare Collaborative (IHC) to host Working Together to Improve Diagnosis, a virtual three half-days conference on May 25—27, 2021. We hope you'll join us!

Why focus on the diagnostic process?
An analysis of Constellation medical professional liability (MPL) claims revealed that diagnostic errors are frequent, costly, and preventable. Constellation's MPL claim analysis also found that diagnostic error is the #3 most frequent allegation and #1 most costly. The virtual conference will highlight the alarming impact diagnostic error has on patients, residents, care teams, and organizations. It will energize teams to identify opportunities for improving diagnostic processes in their own organizations by sharing best practices and successes from industry experts. Additionally, it will highlight the importance of communication when a harm event does occur.

> "It wasn't until we looked at our own [Constellation®] claim data, collectively, that we realized just how much of an impact we might make by not only raising awareness and focusing our efforts on diagnostic error, but also how we might be a part of solving some of the contributing factors to its occurrence."
>
> Laurie C. Drill-Mellum, MD, MPH
> Chief Medical Officer,
> Constellation

Keynote Presenters:

- David Newman-Toker, MD, PhD, Director, Armstrong Institute Center for Diagnostic Excellence, Johns Hopkins Medicine
- Suzanne Schrandt, JD, Senior Engagement Advisor, Society to Improve Diagnosis in Medicine
- Thomas H. Gallagher, MD, Department of Medicine, University of Washington

National Presenters:

- Dana Siegal, RN, CPHRM, CPPS, Director of Patient Safety, CRICO Strategies
- Julia Prentice, PhD, Research Director, Betsy Lehman Center for Patient Safety
- Timothy McDonald, MD, JD, Chief Patient Safety and Risk Officer at RLDatix, Institute for Quality and Safety and Professor of Law, Loyola University
- Bruce Lambert, PhD, Professor, Department of Communication Studies, and Director, Center for Communication and Health, Northwestern University

Who should attend? Executive leaders, physicians, administrators, patient safety/risk personnel, system engineers, quality leaders, and practice leaders from hospitals, clinics, senior living, and long-term care organizations. Due to financial constraints on healthcare organizations in the COVID-19 environment, *Constellation is waiving the registration fee for the conference.* Do you see a session that piques your interest? If you're unable to attend the full, three half-days event, we welcome you to register for a single half-day or for two half-days.

LEARN MORE & REGISTER

SOCIETY to IMPROVE DIAGNOSIS in MEDICINE

In addition, conference participants will have the opportunity to join a new Collaborative dedicated to improving the diagnostic process—a year-long, virtual quality improvement group with a focus on reducing harm events and the subsequent financial burdens while supporting the implementation of new strategies and shared experiences. Learn more about the Collaborative.

About Constellation®
Constellation is a growing portfolio of medical professional liability insurance and partner companies working *Together for the common good.®* Formed in response to the ever-changing realities of health care, Constellation is dedicated to reducing risk and supporting physicians and care teams, thereby improving business results.

About Iowa Healthcare Collaborative
The Iowa Healthcare Collaborative (IHC) is a provider-led, patient-focused, nonprofit organization dedicated to sustainable healthcare transformation. Nationally recognized for achieving demonstrable and sustainable improvements across healthcare settings and disciplines, IHC placed those that deliver care in a leadership position to drive improvements and accelerate change. This mission is possible because of IHC's unified approach to healthcare delivery and strong vision for change.

New Implementation Guide Focuses on Transforming Education of Diagnostic Reasoning

The Society to Improve Diagnosis in Medicine (SIDM) has partnered with the Southern California Permanente Medical Group (SCPMG) and the Human Diagnosis Project (Human Dx) through a grant from the Coverys Foundation to create an education intervention to improve practicing physicians' diagnostic reasoning. The newly developed implementation guide, *Transforming Education on Diagnostic Reasoning: Ready, Set, Go!*, describes how to implement a diagnostic reasoning quality improvement intervention that is scalable and can be carried out in any healthcare system.

Why diagnostic reasoning?

Diagnosis is one of the most complex challenges clinicians face. An estimated 40,00 to 80,000 people die each year from diagnostic failures in U.S. hospitals alone. Diagnostic errors that arise through cognitive errors are often associated with faulty perception, failed heuristics, and biases. Clinicians rely on these shortcuts in reasoning to minimize delay, cost, cognitive load, and anxiety in their clinical decision making. In an analysis of a large medical malpractice claims database, failures in clinical judgment were the leading identified cause of serious misdiagnosis-related harms.

While resources aimed towards increasing awareness and reducing the rate of diagnostic errors are available for physicians in training, few collaborative educational programs are available for practicing clinicians.

With this guide, leaders are shown ways to define the problem and establish a goal, prepare communication plans and curate educational materials, and ultimately measure the effectiveness of the intervention.

“The implementation guide offers practical guidance on how to implement a diagnostic reasoning improvement intervention for busy clinicians,” says Gerry Castro, Ph.D., MPH, PMP, SIDM’s Director of Quality Improvement, “Our partners at SCPMG and Human Dx contributed not only technical expertise but real-world experience to the content of the implantation guide”

Who should use this guide?

Any person working in a healthcare system or clinical setting who is looking to implement an innovative educational strategy to improve diagnostic reasoning, including but not limited to:

- Chief Executive Officers, Chief Medical Officers, Chief Operating Officers, Quality Improvement and Patient Safety Leaders, Clinical Department Directors, Medical Education Leaders, Medical Practice Group Executives, and other such senior leaders

Diagnostic reasoning challenge

During the project, SIDM, Human Dx, and SCPMG aimed to assess outcomes in clinical reasoning, diagnostic accuracy, and collaboration in an innovative virtual educational format.

Human Dx provided Global Morning Report (GMR) clinical cases focused on diagnostic reasoning in the following three areas: infectious diseases, cardiology, and cancer. SCPMG recruited participants and administered the educational platform, materials, and administrative support throughout the quality intervention, while SIDM provided consultative support for the project.

The project consisted of physicians from three Kaiser Permanente regions, who participated in solving a set of Human Dx/GMR cases, engaging in a 6-week intervention phase, and again asked to solve a set of Human Dx/GMR cases differing from the first.

Funding for Human Dx and SIDM was provided through a grant from the Coverys Community Healthcare Foundation