# The Struggle over Artificial Intelligence in Healthcare

An STS Research Paper
presented to the faculty of the
School of Engineering and Applied Science
University of Virginia

by

Srikar Chittari

May 9, 2023

On my honor as a University student, I have neither given nor received unauthorized aid on this assignment as defined by the Honor Guidelines for Thesis-Related Assignments.

*Srikar Chittari*

STS Advisor: Peter Norton

As artificial intelligence (AI) proliferates, the healthcare sector has been applying it to mirror expert human cognition. AI may revolutionize health services by diagnosing diseases, assisting patients virtually, personalizing healthcare plans, and making treatment more accessible and effective. Medical AI algorithms collect vast data from around the world and use it to develop and train AI models to predict future cases. Since the data-guided analytics that AI algorithms use are often misrepresented as "data driven," the models are often misperceived as impartial. However, even an AI model trained exclusively on accurate data cannot make decisions that matter to humans without bias. It can either make decisions that encode human biases, for example in the selection and weighting of relevant data, or it can forgo decision making and instead better inform human-made decisions that inevitably remain, as human decisions, subject to human biases.

An AI model may be biased for various reasons. Feeding the model unrepresentative training data is a common point of failure. For example, in the field of genomics and genetics, Caucasians comprise an estimated 80 percent of collected data, meaning healthcare may be unequally distributed and inaccurate (Igoe, 2021). However, employing even representative datasets may produce discriminatory models. This is because patient data, however inclusive, is a reflection of biases we currently have in our healthcare sector. Ashish Jha, director of the Harvard Global Health Institute, revealed "healthcare systems disproportionately mismanage and mistreat black patients and other people of color," and this data – which has decades of systemic healthcare bias baked into it – is being used to train medical AI (Gawronski, 2019). As a result, AI services tend to reinforce discriminatory healthcare treatment to these marginalized groups, making biases more pervasive and systematic.

In the U.S., advocates of healthcare equity have resisted discriminatory biases encoded in medical AI. Advocacies protecting marginalized patients call for improved federal regulations on medical AI development while advocates of transparency in AI development propose strategies that promote accountability within the AI community. Divergent responses to AI in healthcare reflect divergent assessments of the risks it poses. To some, medical AI is a fundamental threat to patients' rights that exacerbates decades of systemic biases; such groups tend to favor restricting medical applications of AI to a great degree. To others, however, medical AI is a promising tool that presents an opportunity to evaluate and alleviate inequitable practices in healthcare.

**Review of Research**

In 2019, a study by Obermeyer et. al. found racial bias in Optum's widely used medical algorithm. When Obermeyer and his colleagues ran routine statistical checks on medical data, they were "surprised to find that people who self-identified as black were generally assigned lower risk scores than equally sick white people" (Ledford, 2019). Obermeyer explained that fixing algorithmic bias is not straightforward, stating it "is not a problem with one algorithm, or one company — it's a problem with how our entire system approaches this problem" (Ledford, 2019). This study has been foundational and widely cited in the literature and has since exposed algorithmic biases in medical AI and magnified systemic biases in healthcare. Current research looks to apply Obermeyer's methodology to the development processes of all medical AI, calling for regular audits and tests to discover biases (Ledford, 2019). Researchers also argue that a team composed of only AI experts may be insufficient and propose diversifying teams developing AI to better serve the diverse patients that their algorithms affect. Ethicists and philosophers need to join AI developers and healthcare experts in the conversation, all of whom "need to be deeply

involved in the development of such technologies from the beginning" (Keskinbora, 2019). A significant portion of the literature calls on the U.S. government to expand federal oversight on AI development, specifically enforcing audits on medical AI and diversifying teams developing clinical algorithms.

In response to Obermeyer's study and calls to action, American policymakers have enacted legislation to increase oversight on AI development, but researchers claim that current regulations are not yet advanced to mitigate certain biases. Recent legislation "protects certain variables by removing fields that would lead to unfair judgment, like race, gender, socioeconomic background, and disability," but these attributes "are required in these health care algorithms so that these groups may receive proper care" thereby making the omissions ineffective (Igoe, 2021). Also, researchers in the literature emphasize that bias can enter at any point in the AI lifecycle from model development to validation to implementation to maintenance and updates in the post-implementation period (Thomasian, 2021). They argue that current regulations lack adequate oversight in the post-implementation period, and policymakers must understand that "bias mitigation should not end with AI model development but, rather, extend across the product lifecycle" (Thomasain, 2021).

Holistically, the literature highlights a need for federal regulations on AI development and deems current regulations inadequate. Most researchers prioritize formulating amendments to these regulations and lament the lengthy turnaround period to enforce changes. Some advocates promote strategies that may enforce accountability in the AI community. For example, federal regulations whose primary concern is not supervising AI development have been exercised to protect marginalized patients from biased AI models. Furthermore, advocates have held AI corporations accountable for their discriminatory algorithms and created incentives to

preemptively uncover biases in their models before they are released (Igoe, 2021). Since these methods may not be long-term solutions, they are often overlooked in the literature. However, these strategies may be valuable in mitigating algorithmic biases until effective federal regulations are implemented and serving as auxiliary methods alongside improved regulations.

**Legal Action to Protect Marginalized Patients**

Advocacies have filed lawsuits to investigate discriminatory AI models and protect patient rights. For example, Arkansas residents with cerebral palsy experienced extreme budget cuts to in-home care, resulting in hospitalizations and disruptions to their lives. With the aid of a local American Civil Liberties Union (ACLU) branch and a lawsuit, errors in the algorithm's characterization of patients with disabilities were found to be responsible (Grant, 2022). As demonstrated by ACLU's intervention to aid the residents of Arkansas, the advocacy group engages in strategic lawsuits to expose unethical AI and protect underrepresented patients in the health sector. ACLU also aims to use legal action to initiate "policy changes and collaboration among key stakeholders, including state and federal regulators, medical, public health, and clinical advocacy groups and organizations… to address these gaps and inefficiencies" in medical AI (Grant, 2022). ACLU continues to pursue legal actions against biased AI models and hopes this will develop coalition building across the healthcare industry to hold discriminatory AI models accountable.

Federal agencies that do not directly oversee AI development have also held AI creators accountable. In 2021, the Federal Trade Commission (FTC) issued a warning to corporations that using biased AI technology may violate consumer protection laws (Landi, 2021). The Fair Credit Reporting Act (FCRA) prohibits companies from releasing consumer records to anyone without

a legitimate reason to have it and denying services to people. When this law was established in 1970, people were not discussing AI; "only in more recent years has the FTC begun to apply language and tools to AI and companies" (Ajao, 2021). Also, the Equal Credit Opportunity Act prevents companies from discriminating on the basis of race, color, religion, national origin, sex, marital status, age, giving the FTC authority to prohibit AI models that discriminate against such factors (Landi, 2021).

In 2018, the FTC took action against RealPage, a real estate software company which "wrongly matched some applicants to criminal records that did not belong to them," denying them housing and other opportunities (Ajao, 2021). The FTC "claimed RealPage didn't take proper steps to ensure the data it provided to landlords and property managers was correct," and the "company paid the federal agency $3 million to settle the charge" (Ajao, 2021). Although the FTC can mitigate AI bias to an extent, "FCRA is definitely not a comprehensive AI regulatory framework," according to John Davisson, a senior counsel at the Electronic Privacy Information Center (Ajao, 2021). He clarifies that these regulations are reactionary, and they do not impose "nondiscrimination requirements" or require "companies that use AI to validate the tools that they're using" (Ajao, 2021). Advocates argue adapting existing federal regulations to catch AI malpractice may be an effective method until improved legislation is implemented.

**Voluntary Ethics Pledges to Prevent AI Discrimination**

While some advocacy groups pursue legal action to protect patient rights, other groups, such as the Algorithmic Justice League (AJL), advocate for increased transparency in the AI development process. In 2018, Joy Boulamwini, the founder of the AJL, launched the Safe Face Pledge to "prevent the lethal use and mitigate abuse of facial analysis and recognition

technology" (Boulamwini, 2019). The initiative promoted transparency during AI development to prevent the weaponization of AI and harmful discrimination. The pledge targeted facial recognition technology (FRT) in law enforcement but had secondary interests in medical AI applications. Although over 40 organizations pledged, the more prolific providers of FRT such as IBM, Microsoft, Google, and Facebook failed to sign on. In 2021, Boulamwini released a statement announcing the end of the Safe Face Pledge initiative.

While Boulamwini deemed the effort successful in "challeng[ing] companies to make actionable, measurable commitments beyond stating their AI ethics principles," she concluded that the pledge fell short of convincing large corporations to change their practices (Boulamwini, 2021). The founder commented that "through their refusal to sign the Safe Face Pledge, industry leaders conclusively demonstrated that self-regulation is not enough to compel the comprehensive mitigation of abuses and lawless uses of FRTs" (Boulamwini, 2021). The AJL founder calls for public education and mobilization as the next steps to raise awareness for the harmful AI practices. Like the ACLU, the AJL calls on AI companies to improve their development practices, but the AJL used voluntary pledges which ultimately failed to influence industry leaders.

EqualAI, a nonprofit organization founded in 2018 to reduce unconscious bias in AI, also has a voluntary pledge to limit algorithmic bias. Their pledge is paired with a framework and checklist to guide companies on how to best test their AI and identify biases. Like the AJL, EqualAI has not yet attracted leading companies to sign their pledge, but their organization remains active and believes more comprehensive guidelines may help raise awareness (Tague, 2022). In 2022, the organization's president, Miriam Vogel, was appointed as Chair of the National Artificial Intelligence Advisory Committee (NAIAC), which advises the U.S. President

and National AI Initiative Office on AI related issues. Vogel looks to apply his methodology of voluntary pledges to national initiatives and promote self-regulation where it is appropriate (Tague, 2022). While the AJL looks to educate and mobilize on a grassroots level, Vogel seeks to enforce EqualAI's ideology nationally.

**Incentives and AI Audits to Preemptively Expose Biases**

Medical AI has drawn the attention of academics specializing in healthcare inequity. Researchers from the Harvard School of Public Health (HSPH) also point to a lack of transparency in AI development as the root cause of bias but propose a course of action differing from the ACLU and the AJL. Trishan Panch, the president-elect of the HSPM Alumni Association, asserts that "there will probably always be some amount of bias," so incentives can be implemented to expose biases (Igoe, 2021). For example, "if researchers or other professionals can prove that data analysis is biased, they can utilize legislation via class action lawsuits," and "this will incentivize private companies to change or to preemptively look at bias before this occurs" (Igoe, 2021). The approach HSPH proposes relies on installing a financial or legal reward system for third parties to discover AI biases, contrasting the methods outlined by other advocacies pursuing legal action and voluntary corporate pledges.

Volunteers from software companies and researchers in academia have begun to organize competitions that challenge developers to test their AI models on diverse data, allowing them to methodically discover biases in their systems. In what is dubbed as a 'bias bounty' competition, participants were tasked to build an AI image detection model that classified each image with "its skin tone, perceived gender, and age group, [making] it easier to measure and spot biases in datasets" (Heikkila, 2022). Microsoft, Amazon, and startup Robust Intelligence committed prize

money for the winners, demonstrating incentives as outlined by researchers at HSPH. Similar competitions have been conducted: Twitter launched the first AI bias bounty in 2021 and Stanford organized its first AI audit challenge. Such competitions have been "hailed by regulators and AI ethics experts" and recommended by the US National Institute of Standards and Technology as the "gold standard," showing AI auditing as an emerging industry (Heikkila, 2022). Auditing competitions reflect the values held by researchers at HSPH: a self-regulated reward system can promote accountability in the AI community.

While corporations such as Mozilla have created tools for AI auditors, there is a shortage of independent contractors to meet the incoming demand for AI audits (Heikkila, 2022). AI researchers hope bias bounty competitions "lead more engineers, researchers, and experts to develop the skills and experience to carry out these audits" and "create a new sector of experts who specialize in auditing AI" (Heikkila, 2022). A reward system with prize money may provide short-term incentives for AI developers and researchers to audit their models, however these engineers must develop auditing skills of their own to cultivate a culture in the AI community that values self-regulation.

**Responses to Strategies Mitigating AI Bias**

*Healthcare Companies Recover Confidence in AI Services*

Employers of medical AI, such as Optum, contest research on algorithmic bias, deflecting the healthcare inequity to other causes. In 2019, Obermeyer and his colleagues worked without salary to identify algorithmic bias in Optum's medical AI, modified the algorithm accordingly, and found that the changes "reduced bias by 84%" (Ledford, 2019). While Optum said by statement that they "appreciate the researchers' work, including their validation that the cost

model was highly predictive of cost," they stated that "the conclusions of the researchers are misleading" (Morse, 2019). The healthcare company also specified that "the cost model is just one of many data elements intended to be used to select patients for clinical engagement programs," and in an earlier statement clarified that "these tools should never be viewed as a substitute for a doctor's expertise and knowledge of their patients' individual needs" (Morse, 2019). When Optum was made the face of algorithmic bias in the healthcare sector, the company seeked to recover confidence in their AI services by acknowledging the general limitations of AI decision-making but not the systemic biases of their industry manifesting in their services.

Since Obermeyer's study, Optum has been conducting annual surveys, polling "500 senior health care executives from leading hospitals, health plans, life sciences companies and employers" to boost client confidence in AI services (Optum, 2023). Their fourth annual survey reported that "99% agree that AI can be trusted for use in health care," displaying to their patients a virtually unanimous belief in medical AI (Optum, 2023). Optum also found that "94% [of health executives] agree that they have a duty to ensure responsible use of AI," alluding that executives at Optum are monitoring AI services to prevent another instance of discriminatory AI service (Optum, 2023). Optum has publicly downplayed research on algorithmic biases and has bolstered its clients' confidence in AI through the use of surveys.

*Expanded Federal Oversight on AI Development*

With growing calls to reform AI regulation, the U.S. Food and Drug Administration (FDA) has expanded its oversight of medical AI to verify training sets and mitigate algorithmic bias. In 2021, the agency employed a new framework that would classify medical AI as "Software as a Medical Device" and "enable the FDA and manufacturers to evaluate and monitor

a software product from its pre-market development to post-market performance" (CDRH, 2021). The creation of this framework was a response to activists who criticized the lack of federal oversight in the post-implementation period of medical AI. Furthermore, the key improvement in this plan was anticipating modifications during a medical product's lifetime and "implement[ing] those changes in a controlled manner that manages risks to patients" (CDRH, 2021). The FDA's concern with post-implementation risks demonstrated that the policy makers had now better understood the evolutionary nature of AI models and the need for constant supervision. As a result, the "FDA would expect a commitment from manufacturers on transparency and real-world performance monitoring… as well as periodic updates to the FDA on what changes were implemented" (CDRH, 2021).

Although the FDA's recent framework proposed oversight over a greater portion of the product's lifecycle, criticism stemmed from which AI devices were classified as 'Software as Medical Devices.' Due to loopholes in the regulations, "a large proportion of artificial intelligence algorithms are exempted by this definition and are already in widespread use throughout the health sector" (Thomasian, 2021). While the FDA proposes a "much-needed precedent for fairness in medical informatics, there is no guarantee that such a framework would trickle down to the FDA-exempt majority of healthcare AI" (Thomasian, 2021). The establishment of periodic federal guidance for medical AI has addressed some of the previous inadequacies, but the FDA cannot expect regulations to be effective if they only apply to a small portion of AI products in the healthcare sector.

**Conclusion**

Until effective federal legislation is enacted, advocates must continue to engage directly with the AI creators to minimize biases in medical AI. These methods may develop accountability amongst AI developers and encourage them to evaluate algorithmic models through the lens of inequity present in healthcare. In the long-term, this self-regulating mindset may manifest in a new industry sector to audit AI models, which would be valuable in instances where federal regulations fail to protect consumers from discriminatory products. As AI continues to serve industries, advocates of AI equity must hold developers to a high standard in discussing systemic biases with subject specialists, ethicists, and philosophers before creating tools that perpetuate existing biases. Furthermore, champions of AI may find it worthwhile to clarify misconceptions about the impartiality of AI models to prevent developers and consumers alike from interpreting AI-made decisions as objective truths. Consumers must understand that AI, after all, is a tool created by society and trained by data on society's people, actions, ideas, and opinions.

**References**

Ajao, E. (2021, Oct. 19). FTC pursues AI regulation, bans biased algorithms. *TechTarget*. www.techtarget.com/searchenterpriseai/feature/FTC-pursues-AI-regulation-bans-biased-algorithms

Buolamwini, J. (2021, Feb. 8). Announcing the sunset of the Safe Face Pledge. *Medium*. medium.com/@Joy.Buolamwini/announcing-the-sunset-of-the-safe-face-pledge-36e6ea9e0dc5

Buolamwini, J. (2019, Feb. 7). Artificial Intelligence Has a Problem With Gender and Racial Bias. Here's How to Solve It. *Time*. time.com/5520558/artificial-intelligence-racial-gender-bias

CDRH (2021, Sep. 22). Center for Devices and Radiological Health. Artificial Intelligence and machine learning in software. U.S. Food and Drug Administration.

www.fda.gov/medical-devices/software-medical-device-samd/artificial-intelligence-and-machine-learning-software-medical-device

Gawronski, Q. (2019, Nov. 6). Racial bias found in widely used health care algorithm. *NBC News*. www.nbcnews.com/news/nbcblk/racial-bias-found-widely-used-health-care-algorithm-n1076436

Grant, C. (2022, Oct. 3). Algorithms Are Making Decisions About Health Care, Which May Only Worsen Medical Racism. American Civil Liberties Union. www.aclu.org/news/privacy-technology/algorithms-in-health-care-may-worsen-medical-racism

Heikkila, M. (2022, Oct. 20). A bias bounty for AI will help to catch unfair algorithms faster. *MIT Technology Review*. https://www.technologyreview.com/2022/10/20/1061977/ai-bias-bounty-help-catch-unfair-algorithms-faster/

Igoe, K. (2021, March 12). Algorithmic Bias in Health Care Exacerbates Social Inequities — How to Prevent It. Harvard School of Public Health. www.hsph.harvard.edu/ecpe/how-to-prevent-algorithmic-bias-in-health-care

Landi, H. (2021, April 22). FTC issues warning that using biased AI could violate consumer protection laws. *Fierce Healthcare*. www.fiercehealthcare.com/tech/ftc-issues-warning-using-biased-ai-could-violate-consumer-protection-laws

Ledford, H. (2019, Oct. 24). Millions of black people affected by racial bias in health-care algorithms. *Nature*. www.nature.com/articles/d41586-019-03228-6

Morse, S. (2019, Oct. 25). Study finds racial bias in Optum algorithm. *Healthcare Finance*. www.healthcarefinancenews.com/news/study-finds-racial-bias-optum-algorithm

Optum (2023). Responsible use of artificial intelligence. www.optum.com/business/insights/health-care-delivery/page.hub.responsible-artificial-intelligence.html

Tague, M. (2022, Aug. 2). Business leaders commemorate anniversary of EqualAI and its new leadership role on the National Artificial Intelligence Advisory Committee. LivePerson. pr.liveperson.com/2022-08-02-Business-leaders-commemorate-anniversary-of-EqualAI-and-its-new-leadership-role-on-the-National-Artificial-Intelligence-Advisory-Committee

Thomasian, N.M.; Eickhoff, C.; and Adashi, E.Y. (2021, Nov. 22). Advancing health equity with artificial intelligence. *Journal of Public Health Policy 42*, 602–611. Web of Science