



How Implicit Bias Affects AI in Healthcare

Tania M. Martin-Mercado, PhD^{1,2*} and Damon House³

¹Research and Development, Phronetik, Inc., Flower Mound, TX, United States

²Customer Transformation and Innovation, Microsoft Corporation, Redmond, WA, United States

³Public Sector, Microsoft Corporation, Redmond, WA, United States

*Corresponding author: Tania Martin-Mercado, Research and Development, Phronetik, Inc., Flower Mound, TX, United States, E-mail: hello@drtaniam.com

Received date: April 1, 2022, Manuscript No. JDTBA-22-59978;

Editor Assigned date: April 4, 2022, PreQC No. JDTBA-22-59978 (PQ);

Reviewed date: April 19, 2022, QC No. JDTBA-22-59978;

Revised date: April 25, 2022, Manuscript No. JDTBA-22-59978 (R);

Published date: April 29, 2022, DOI: 10.4172/jdtba.1000256

Abstract

Implicit bias is a natural survival instinct inherent in human beings. However, when implicit bias is not addressed in a healthcare setting, health disparities, gaps in care, and discrimination can occur in a clinical setting. Algorithmic bias in healthcare still exists today, and as new technologies and digital innovations are introduced into the care continuum, leaders must acknowledge that many of the foundations the tools are built upon contain damaging bias that negatively impacts patient care. Taking deliberate steps to mitigate these biases can reduce health disparities and allow the promise of artificial intelligence to be fully realized in improving patient outcomes.

Keywords: Implicit bias; AI: Artificial Intelligence; Minority health; Health disparities; Bias; Decision support; Diagnostic techniques

Introduction

What is Implicit Bias?

Implicit bias is a form of bias that occurs automatically and unintentionally, that nevertheless affects judgments, decisions, and behaviors [1]. It occurs in everyday life using existing pathways and shortcuts to filter out the incessant stream of information from the environment. From decisions about who to trust to deciding what to believe, the human brain has been trained by generations of interactions to wire these shortcuts to simplify our lives. Sometimes, these shortcuts are objective and harmless to others; however, there are other shortcuts that are based on preconceived notions that may have a detrimental impact on interactions with the surrounding world.

Ultimately, implicit bias is a natural instinct. Bias is not evil; it stems from basic survival instinct. It is human and feeds the natural sense to belong [2]. The neural circuits that govern social behavior and reward arose early in vertebrate evolution and are present in birds, reptiles, bony fishes, and amphibians, as well as mammals. While there is little information on reward pathway activity in humans during

in-group versus out-group social situations, there are some tantalizing results from studies on other mammals.

Calling attention to implicit bias, particularly when conscious decisions are being made in relation to healthcare decision making and decision support, should be a welcome conversation with an aim to improve patient care and the patient experience.

Impact of Bias on Decision Making in Healthcare

Implicit bias can and does shape the provider-patient interaction. Biases are inserted into the clinical engagement by a provider or patient based on mistrust or even a fundamental misunderstanding of the clinical circumstances. At times, they are programmed into the very processes and procedures that the provider's organization or a regulatory body have suggested or even mandated for a particular diagnosis. Healthcare organizations need to dive deeper into the implications of bias for patient treatment at both an individual and institutional level. Leaders in healthcare IT and informatics need to be involved in these conversations and solutions at the highest levels.

Bias in Clinical Care Algorithms

It is critically important to acknowledge that the idea of race is a social construct, not a biological one, nor is it a reliable measure of genetic differences. Yet, using race as a factor is commonplace when designing clinical algorithms. Examples of this issue that are occurring in real-time across the healthcare landscape in 2022 include:

- “Black man’s cocktail”: This is based on a pre-determined set of health-related issues that are common among African American males, including diabetes, hypertension, and other factors. These assumptions and biases exist before the point of care. They are not a SOAP note, nor are they limited by intake. There is no objectivity in this diagnosis; it is simply a subjective diagnosis with a resulting set of prescriptions and clinical directions that are meant to save time and money rather than accurately account for the unique needs of the individual patient.
- Race correction: This issue raises the threshold for use of clinical resources among minority patients. Levels of a patient’s insurance may be the prevailing determinant of the level of care that is prescribed. When the high correlation between a patient’s race and their income is examined, it is easy to see how race drives this bias. Race correction is currently built into the tools used today in clinical settings to diagnose and determine treatment for patients [3].
- Kidney failure: Creatinine levels in a patient’s blood are commonly used as an indicator of proper kidney function; the less creatinine a patient’s blood has, the better the patient’s kidney functions. African American patients’ creatinine levels are commonly adjusted under the assumption that these patients have more muscle mass—a fundamental misdiagnosis that doesn’t account for the individual patient’s circumstances. This results in African American patients having higher rates of end-stage renal disease than their Caucasian counterparts [4].
- Breast cancer: An online tool that estimates the risk of breast cancer calculates a lower risk for African American or Latinx women than their Caucasian counterparts, even when every other risk factor is identical. This typically deters minority women from undergoing the necessary screening that is critical for diagnosing this issue and getting them the treatment that they need early enough to improve their outcomes.

Strategies to Address Bias and Increase Data Fairness

Data science teams must investigate and identify ways to put action behind the initiative. The greatest opportunity to drive effective solutions is to bring diverse perspectives to solve the problem.

Concentrated effort to hire diverse data teams is a necessity-it is critically important to have individuals on a data team that resonate with the data being collected and analyzed. As data science leaders seek to hire diverse team members, several factors must be considered.

- Inspire creativity - Individuals with diverse experiences and perspectives offer alternative outlooks and solutions
- Support innovative ideas - Diverse backgrounds lead to new ideas and approaches to problem solving
- Advocate for cultural awareness - Patients come in all shapes, sizes, colors, and backgrounds. Awareness of this diversity is increased when teams are themselves diverse
- Pinpoint blind spots - Allow others to address implicit bias where others may not see it
- Improve team relationships - Diverse teams often have better and more productive collaboration
- Foster empathy and compassion - Diverse teams create an atmosphere where understanding differences, ideas, and perspectives contribute to empathy for others

And lastly, when the argument of perceived difficulty in finding diverse, qualified members to include on these teams, leaders must push through to conversations about better training, development of employment pipelines, and old-fashioned hard work to find the necessary team members.

Healthcare leaders must leverage better tools to reduce bias-as an industry, normalizing the use of available tools to reduce bias addresses the problem at the core. Some of the current tools used by data scientists to mitigate bias include:

- What-If tool - Released by Google in 2018 as part of its People +AI research initiative
- AI fairness 360 - An open-source toolkit of metrics to check for unwanted bias in datasets, ML models and algorithms
- TCAV - Research initiative called "testing with concept activation" released by Google to detect bias in ML models
- Skater - An oracle initiative; Python library for complex or black-box model that uses various techniques to detect bias by understanding how a model makes a prediction based on the data it receives

These tools do not represent an exhaustive list -They can help to advance the necessary conversations and examine the status quo. During this analysis, the following questions must be asked: "Who is evaluating the results of an audit?" "Is the data science team diverse enough to connect with the data being collected, analyzed, and scrutinized for bias, or is it a homogenous team?" Without diversity in the teams addressing these concerns, the reviews may end up reinforcing the very biases sought to be eliminated.

Accountability must be encouraged to demonstrate an ongoing commitment to addressing bias. Stakeholders should offer support to provider team members to speak up, even anonymously, when the data seems unfair or biased. The principle of dignity over surveillance creates positive conditions where bias can be addressed without guilt or applying defensiveness or emotion to the issue. The idea of a fairness focus fosters accountability, though it must not confuse sameness with fairness. Prioritizing fairness in an accountability process establishes connections and promotes diverse opinions.

A focus on correction rather than blame reminds individuals that the goal is to address and fix the biases rather than finding scapegoats for the situation and assessing blame.

Artificial Intelligence (AI tools are increasingly used to determine who gets healthcare and can unintentionally increase existing racial bias in medicine. The challenge is to address and correct the lack of inclusion of AI among developers, researchers, and funders in tools. Our charge is to incorporate equity in algorithmic design. By centering on health equity and racial justice, AI tools can break down, rather than enhance, structural inequities. The mandate is to embrace more diverse and complete data sets. Medical data sharing should become normalized to consolidate large, diverse data sets required to train algorithms.

Action Plan to Address Ethical Data Concerns

To properly address these concerns, several critical steps must be taken. Begin with conducting a premortem. A premortem can identify potential biases before they happen. Team members should be encouraged to be open about any misgivings they have, any potential for bias, and any ethical considerations. Ethical considerations can include how the data is collected, from whom, and which data points are and are not necessary to solve the problem. Next, address excluded or overrepresented factors in the dataset. There are social, cultural, and economic factors reflected in a healthcare dataset. Any bias can create unintended unfairness and ethical concerns. Data teams need to ask themselves if they will leave out some groups of the population who are perceived as worse off because of the algorithm's design or its possible consequences. And lastly, design questions for bias impact. There are templates available for data science teams to evaluate bias impact [5]. This is a low-cost, self-regulated way to define scope and predict bias during premortem. The use of design questions for bias impact can filter out potential bias with a discrete set of questions. There are sample bias detection templates available online with a set of questions that begin a premortem conversation [1].

Conclusion

By addressing the lack of inclusion among developers, researchers, and funders of artificial intelligence tools and solutions, bias can be further mitigated and potentially excluded from the tools used in healthcare today. Data science and technology leaders in healthcare can work closely with clinical leaders to incorporate equity in algorithmic design and normalize collecting more diverse and complete datasets required to train algorithms.

Careful attention to the diverse makeup of data science teams in healthcare and clinical settings, while providing a culture of accountability, will further reduce and mitigate the damage caused by algorithmic bias. It is time to move beyond the rhetoric and written promises and offer an actionable organizational model to frame best practices around reducing implicit bias in AI tools used in a healthcare setting.

Acknowledgements

Thank you to Phronetik, Microsoft, Claude Louis-Charles, Dr. Desmond Stubbs, and Marvin A Martin for useful discussions in support of this important topic.

References

1. National Institutes of Health OoWD (2022) Implicit Bias.
2. Psychology Today. Bias.
3. Vyas DA, Eisenstein LG, Jones DS (2020) Hidden in plain sight- Reconsidering the use of race correction in clinical algorithms. *N Engl J Med* 383: 874-882.
4. Tsai J (2021) Jordan Crowley would be in line for a kidney-If he were deemed white enough. *Slate*.
5. Lee NT, Resnick P, Barton G (2019) Algorithmic bias detection and mitigation: Best practices and policies to reduce consumer harms. *Brookings*.