Intersect, Vol 19, No 1 (2025)

Ethical AI Integration in Critical Healthcare & Mitigating Biases: What Are The Ethical Considerations And Strategies To Mitigate Biases When Integrating AI Powered Decision Making Systems In Critical Healthcare Settings?

Muhammad Moeez Jamil and Abdullah Awan
Syed Babar Ali School of Science and Engineering: LUMS

Abstract

The increasing integration of artificial intelligence (AI) into society has created myriad legal and ethical dilemmas. This article examines the key concerns raised by the adoption of AI, particularly in healthcare settings. It explores issues of privacy and surveillance, bias and discrimination, and the interplay between artificial intelligence and human judgment. In addition, the potential risks associated with new digital technologies will be discussed, including data breaches and inaccuracies. Given the patients' vulnerability in dealing with the doctor and the lack of clearly defined rules in this area, the importance of mitigating errors in treatment protocols was highlighted. The document calls for measures to ensure algorithmic transparency, protect privacy, protect all stakeholders and address cyber vulnerabilities to support the responsible implementation of AI in healthcare and beyond.

Key Words: Artificial Intelligence, Ethical issues, Bias, Healthcare

Introduction

As the healthcare industry harnesses the transformative potential of AI-powered decision-making systems, there is a burning ethical issue: potential for bias inherent in AI algorithms. These algorithms learn from large datasets that unintentionally reflect people's historical biases, thereby perpetuating and exacerbating inequalities in health outcomes, particularly among vulnerable populations. The far-reaching consequences of AI bias are reflected in unequal access to treatments and diagnostics, which threaten patient safety and undermines trust in healthcare systems.

The aim of this research is to examine the ethical considerations related to the introduction of AI-based decision-making systems in critical healthcare settings. Our main goal is to explore effective strategies and best practices to reduce bias. By exploring different dimensions such as algorithmic transparency, accountability and inclusivity, we aim to provide valuable insights into how to support the ethical integration of AI and ultimately improve patient care.

Through this research, we aim to clarify the importance of ensuring ethical AI practices in healthcare. By detecting and proactively removing bias in AI systems, stakeholders including healthcare providers, policymakers and technology leaders can work together to create a more AI-enabled, equitable, trustworthy and patient-centric healthcare landscape. By adopting these principles of responsible AI implementation, we work to improve patient outcomes for and advance healthcare by creating a healthcare ecosystem that reflects the highest ethical standards.

Ethical Considerations in AI-powered Decision-making Systems

Data Collection and Selection

An essential ethical consideration in the integration of AI in healthcare revolves around the issue of bias and the pursuit of fairness. Unintentional biases in algorithms can result in the treatment of particular patient groups differently. To identify possible issues with AI systems, it is essential to have a thorough grasp of the various forms of biases, including data bias, algorithmic bias, and selection bias. [1] To stop the recurrence of healthcare inequities based on characteristics like race, gender, ethnicity, or socioeconomic position, emphasizing fairness in AI-powered decision-making becomes crucial.

Transparency and Explainability

In discussions surrounding artificial intelligence law, the problem of algorithmic openness has been a prominent concern. The increasing usage of AI in high-risk situations has increased the need for responsible, equitable, and open AI design and governance. Achieving accessibility and understandability of information is one of transparency’s key components. However, details about algorithm functionality are frequently purposefully concealed or rendered difficult to ascertain. [2] Often referred to as the "black box" problem, [3] the lack of transparency of AI systems can affect trust and accountability. Critical healthcare environments require explainable AI models to ensure healthcare professionals and patients understand the rationale for AI-enabled decisions. Transparency in the devel-opment of AI algorithms and models is essential to spot and eliminate potential errors and hold AI systems accountable for their actions.

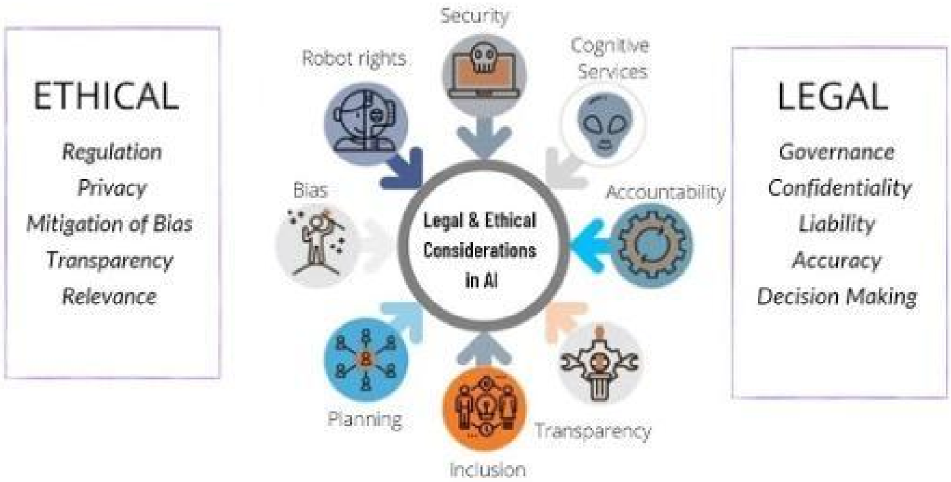


FIGURE 1. The utilization of artificial intelligence in healthcare presents a range of ethical and legal dilemmas.

Accountability and Responsibility

As AI becomes more prevalent in critical healthcare, ensuring accountability for AI decisions is paramount. Healthcare providers, developers, and stakeholders need to understand their role in AI decision-making and be accountable for outcomes. Ethical guidelines and legal frameworks can provide a basis for assigning responsibilities and dealing with any negative consequences of AI decisions. Selection bias in the datasets used to create AI algorithms is a common problem.

Buolamwini and Gebru's [4] research found a bias in automated facial recognition systems and their datasets, leading to lower accuracy in recognizing Black people, especially women. Machine learning relies on vast amounts of data, often from clinical trial databases, most of which represent specific population groups. Therefore, when these algorithms are applied to underserved and potentially underrepresented patient populations, they are more likely to fail.

Inclusivity and Accessibility

Integrating AI into critical healthcare should prioritize inclusion and accessibility. Ethical AI systems should benefit all patient populations, regardless of background or situation. Particular attention should be paid to vulnerable groups to avoid exacerbating existing health inequalities and to ensure equal access to health services. This illustrates the diverse ethical and legal considerations linked to the adoption of AI in healthcare settings.

Recognizing and Addressing Biases in AI-driven Decision-making Systems

Data is the foundation of AI systems, and skewed data can lead to skewed results. Ethical data collection practices include minimizing bias in data collection, identifying and eliminating historical bias in datasets, and ensuring diverse representation to avoid excluding underrepresented groups.

Racial Biases

The US healthcare system relies on commercial algorithms to make healthcare decisions. A study by Obermeyer et al. [6] reveals evidence of racial bias in a widely used algorithm. Black patients assigned the same risk level by the algorithm are sicker than White patients with the same risk level. This racial bias is leading to a significant decrease in the number of Black patients being identified for secondary care. The bias in the algorithm is due to the use of healthcare costs as an indicator of healthcare needs. Because less money is spent on black patients with the same needs, the algorithm incorrectly concludes that black patients are healthier than equally ill white patients. [7]

According to the analysis of a national dataset containing 3,695,943 commercially insured patients confirmed our findings. According to one measure of predictive bias calculated in their dataset, Black patients were found to have 48,772 more active chronic conditions than White patients, even when considering their risk scores. This observation serves as an illustration of how biases can unintentionally arise in the data analysis process. [8]

Additionally, in another study, the JAMA Dermatology Network [9] found that there were discrepancies in skin cancer diagnosis among people of different skin tones. Dermatological models for detecting skin cancer or potentially cancerous plaques are mostly trained on White subjects, which reduces the accuracy of skin cancer diagnosis in Black patients. Although people with darker skin are generally at a lower risk of developing skin cancer, as stated by the American Academy of Dermatology Association, [10] diagnostic challenges in correctly detecting skin cancer in this population remain significant. Skin cancer in people of color is often diagnosed at a later stage, making the treatment process more difficult. The lack of diversity in training AI models makes it difficult for them to quickly and accurately diagnose skin cancer in patients with darker skin. They may have a hard time seeing contrast effectively, or they may even overlook a condition altogether, which can lead to misdiagnosis. [11]

Gender Disparity in Chest X-Rays

Artificial intelligence is making great strides in healthcare, especially in the analysis of medical images such as X-rays and MRI scans. However, these systems can inherit errors from training data and perpetuate and amplify those errors. A 2020 study published in PNAS [12] found that gender imbalance in training data from computer-aided diagnostic (CAD) systems resulted in lower diagnostic accuracy of an underrepresented group. In particular, when the CAD system was trained primarily on male radiographs, it showed significantly lower accuracy in diagnosing females. To increase overall accuracy, AI algorithms should be trained using large and diverse datasets containing balanced information from diverse populations. This approach is key to reducing stigma and ensuring fair and equitable health outcomes for all patients. [12]

Best practices and guidelines for the ethical integration of AI into critical healthcare environments

Although prejudice and assumptions are part of human nature, they must be challenged and eliminated from the healthcare system. Implementing specific actions at your facility can help reduce and eliminate medication errors. Here are some steps you can take to make that happen.

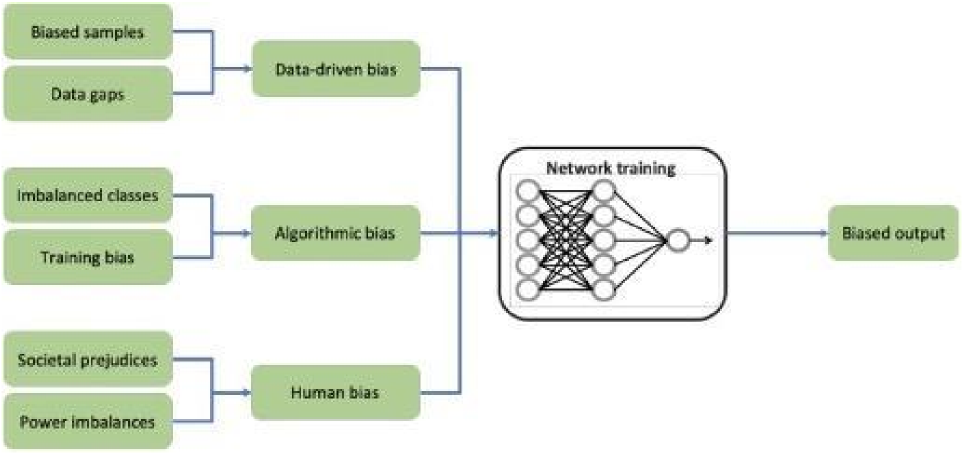
Correct framing of the problem

In order to solve the algorithmic error problem in AI, the hypothesis and the concept of the model must be formulated and the desired results must be clearly defined. A crucial first step is to encourage diversity and representation within the research team. It’s not just about involving clinical experts and data scientists in the process, but also key stakeholders, people from under-represented populations and end-users. [13] A diverse team is essential from the early stages of problem formulation and hypothesis generation to deploying the model in different populations and evaluating its performance, generalizability, and usability. When formulating a problem for a predictive model, it is important to identify the research question, the population of interest, predictors/variables, and the desired outcome. AI developers should proactively consider diversity in model design and hypothesis formulation to avoid unintended biases toward specific patient populations. [14]

Developers must constantly ask themselves whether their algorithm esti-mates could lead to unintended consequences for certain patient groups. To solve this problem, several factors should be considered. First, the configuration of the problem, including the availability of data and the complexity of the idea, requires careful study. Second, it is necessary to involve experts from different fields and multidisciplinary teams with relevant experience. Additionally, considering patient demographics and socioeconomic status, and engaging the community in understanding their diverse experiences and potential biases are critical steps in ensuring a fair and unbiased AI system. [15]

Data Collection and Its Diversity

The different types of errors in prediction algorithms can largely be traced back to the data used in their development. Figure 5.2 An illustration of the various sources of bias that can arise during the training of machine learning algorithms. [16]



Sample bias, a common form of data bias, occurs when data used to develop AI algorithms is collected from patient cohorts that do not accurately represent the entire population to which the system is intended to be applied. [17] AI developers should try not to rely solely on data from a single institution. Instead, they should integrate disparate data sets to ensure that key variables such as race, ethnicity, language, culture, and social determinants of health are comprehensively captured and incorporated into prediction algorithms designed to minimize bias.

Throughout the history of clinical trials, women and minority groups have been underrepresented as study participants, providing evidence that these groups of approved drugs had fewer benefits and more side effects. Symptoms that are commonly associated with common health problems are often based on male experiences, leading to discrepancies in diagnosing health problems in women. For example, while chest pain is generally considered an important indicator of a heart attack, women are more likely to experience symptoms such as dizziness, shortness of breath or nausea. As a result, healthcare professionals may be slower to diagnose women who have had a heart attack, and in some cases may miss the diagnosis. [18]

Guaranteeing transparency, privacy, and appropriate regulatory oversight

Transparency plays a key role in ensuring fairness in AI decision-making. One approach is to compare algorithm decisions made by humans and vice versa to encourage mutual accountability and mitigate bias. To achieve this transparency, the existence of a supporting infrastructure covering technical, regulatory, economic and data protection elements [19] is essential to provide diverse and comprehensive data for training algorithms.

According to Dr. Peter Emb´ı, [20] associate dean for research in informatics and health services at Indiana University School of Medicine, the performance of algorithms can vary depending on data, parameters and human interactions. Continuous assessment is critical to identifying and correcting inherent and systemic inequalities in the health system. Developing tools and skills that enable systematic monitoring and vigilance in the use of algorithms in healthcare is critical to maintaining integrity and preventing unintended harm.

The bias of AI in healthcare often unintentionally mirrors the bias of medical professionals as the AI model learns from the diagnoses of these professionals. So if there are errors in the decision-making process of the medical professional, these will also be visible in the output of the AI algorithm. To ensure AI models are bias-free, physicians and nurses must first address and address their own implicit biases.

Discussion

With increasing evidence of bias in the development and use of AI-powered predictive models, it is paramount to identify and address sources of bias at every stage. However, our current understanding of AI bias only scratches the surface. A concerted effort is needed to harness the full potential of AI technologies to improve healthcare and prevent growing inequalities. AI-powered decision-making systems have the potential to revolutionize critical healthcare environments, but their integration requires careful attention and ethical considerations. For AI systems to improve patient outcomes while meeting ethical standards, they must mitigate bias, ensure transparency and accountability, promote inclusion, and engage stakeholders. Integrating ethical AI can pave the way for a fairer and more efficient healthcare system, building trust and improving patient care. Stakeholders including clinicians, AI researchers, patient advocacy groups, health equity researchers,government agencies and industry players worldwide need to come together to improve representation in AI models. Only by working together can we ensure that AI tools have a positive impact on healthcare and do not perpetuate inequalities.

References

Derraik, J. G. B., Albert, B. B., de Bock, M., Butler, E. M., Hofman, P. L., & Cutfield, W. S. (2018). Socioeconomic status is not associated with health-related quality of life in a group of overweight middle-aged men. PeerJ, 6, e5193. https://doi.org/10.7717/peerj.5193

Safdar, N. M., Banja, J. D., & Meltzer, C. C. (2020). Ethical considerations in artificial intelligence. European Journal of Radiology, 122, 108768. https://doi.org/10.1016/j.ejrad.2019.108768

Yasar, K., & Wigmore, I. (2023). What is black box AI? TechTarget. https://www.techtarget.com/whatis/definition/black-box-AI

Buolamwini, J., & Gebru, T. (2018). Gender shades: Intersectional accuracy disparities in commercial gender classification. In Proceedings of the Conference on Fairness, Accountability and Transparency (pp. 77–91). PMLR.

Naik, N., Hameed, B. M. Z., Shetty, D. K., Swain, D., Shah, M., Paul, R., Aggarwal, K., Ibrahim, S., Patil, V., Smriti, K., Shetty, S., Rai, B. P., Chlosta, P., & Somani, B. K. (2022). Legal and ethical considerations in artificial intelligence in healthcare: Who takes responsibility? Frontiers in Surgery, 9, 862322. https://doi.org/10.3389/fsurg.2022.862322

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

Sklar, J. (2022). Artificial intelligence exacerbates and mitigates racial bias in health care. Journalist's Resource. https://journalistsresource.org/home/research-artificial-intelligence-can-fuel-racial-bias-in-health-care-but-can-mitigate-it-too/

Adamson, A. S., & Smith, A. (2018). Machine learning and health care disparities in dermatology. JAMA Dermatology, 154(11), 1247–1248. https://doi.org/10.1001/jamadermatol.2018.2348

Shue-McGuffin, K. D., & Powers, K. (2022). Skin cancer in people of color. Journal of Dermatology Nurses' Association, 14(4), 152–160. https://doi.org/10.1097/JDN.0000000000000711

Fulmer, J. (2022). Addressing AI and implicit bias in healthcare. TechnologyAdvice. https://technologyadvice.com/blog/healthcare/ai-bias-in-healthcare/

Elul, Y., Rosenberg, A. A., Schuster, A., Bronstein, A. M., & Yaniv, Y. (2021). Meeting the unmet needs of clinicians from AI systems showcased for cardiology with deep-learning–based ECG analysis. Proceedings of the National Academy of Sciences, 118(24), e2020620118. https://doi.org/10.1073/pnas.2020620118

Dankwa-Mullan, I., Scheufele, E. L., Matheny, M. E., Quintana, Y., Chapman, W. W., Jackson, G., & South, B. R. (2021). A proposed framework on integrating health equity and racial justice into the artificial intelligence development lifecycle. Journal of Health Care for the Poor and Underserved, 32(2), 300–317. https://doi.org/10.1353/hpu.2021.0045

Van de Sande, D., Van Genderen, M. E., Smit, J. M., Huiskens, J., Visser, J. J., Veen, R. E., Van Unen, E., Hilgers, O., Gommers, D., & van Bommel, J. (2022). Developing, implementing and governing artificial intelligence in medicine: A step-by-step approach to prevent an artificial intelligence winter. BMJ Health & Care

Informatics, 29(1), e100495. https://doi.org/10.1136/bmjhci-2021-100495

Nazer, L. H., Zatarah, R., Waldrip, S., Ke, J. X. C., Moukheiber, M., Khanna, A. K., Hicklen, R. S., Moukheiber, L., Moukheiber, D., Ma, H., & Mathur, P. (2023). Bias in artificial intelligence algorithms and recommendations for mitigation. PLOS Digital Health, 2(6), e0000278. https://doi.org/10.1371/journal.pdig.0000278

Norori, N., Hu, Q., Aellen, F. M., Faraci, F. D., & Tzovara, A. (2021). Addressing bias in big data and AI for health care: A call for open science. Patterns, 2(10), 100347. https://doi.org/10.1016/j.patter.2021.100347

Vokinger, K. N., Feuerriegel, S., & Kesselheim, A. S. (2021). Mitigating bias in machine learning for medicine. Communications Medicine, 1(1), 25. https://doi.org/10.1038/s43856-021-00025-6

Gaggin, H., & Oseran, A. (2020). Gender differences in cardiovascular disease: Women are less likely to be prescribed certain heart medications. Harvard Health Publishing. https://www.health.harvard.edu/blog/gender-differences-in-cardiovascular-disease-women-are-less-likely-to-be-prescribed-certain-heart-medications-2020071620553

Kaushal, A., Altman, R., & Langlotz, C. (2020). Health care AI systems are biased. Scientific American. https://www.scientificamerican.com/article/health-care-ai-systems-are-biased/

Eliminating bias from healthcare AI critical to improve health equity. (n.d.). EurekAlert! https://www.eurekalert.org/news-releases/630788