Leadership's Role in Preventing Algorithmic Bias in Healthcare

Marywood University

Keishalee Shaw

SLAS 6013: Qualifying Seminar

PhD in Strategic Leadership and Administrative Studies

Fall 2023

Abstract

This research outlines the possible sources of algorithmic bias within healthcare Artificial Intelligence (AI) and Machine Learning (ML) tools and systems, their impacts, and possible bias mitigation mechanisms within the reach of healthcare leadership. Algorithmic bias is a critical issue facing AI-powered programs and systems' adoption within the healthcare sector. While there are numerous advantages to employing AI in healthcare, such as enhancing precision, productivity, and objectivity, there are drawbacks. AI algorithms can perpetuate entrenched societal biases and stereotypes within the United States healthcare system, partly due to the historical legacy of discrimination embedded in the data collected over time, reflecting the conscious and unconscious biases prevalent during the data gathered. Studies have indicated that algorithmic bias, mostly from biased algorithm training data, can perpetuate systemic and structural health disparities. Consequently, the research implies that healthcare leaders must mitigate algorithmic bias during data collection, analysis, and use of artificial intelligence and machine learning systems within healthcare settings. Philosophically, leaders should ensure that any algorithm developed must have adequate medical knowledge based on quality training datasets to eliminate disparate impacts during the adoption and implementation phases. Strategically, healthcare leaders should ensure the development of diverse and inclusive algorithms and model development by leveraging insights from diverse teams hired to identify and address unconscious bias in data and intelligence. Lastly, it is incumbent upon healthcare leaders to equip their teams with ongoing training workshops to ensure sufficient awareness and competencies to mitigate algorithmic biases and associated adverse effects effectively.

Keywords: Algorithmic Bias, Artificial Intelligence (AI), Healthcare, Leadership, Machine Learning (ML)

Table of Contents

Introduction

Access to healthcare is an inherent human right that should be available to all, regardless of race, ethnicity, socioeconomic status, gender, or other demographic characteristics. Unfortunately, the healthcare system continues to grapple with inequities, discrimination, adverse outcomes, and stigma, all of which mainly affect marginalized groups (Sanders, 2023). One increasingly pervasive and insidious factor contributing to these problems is the presence of algorithmic bias in healthcare.

AI bias in healthcare arises when artificial intelligence systems generate unfair or discriminatory outcomes. This bias can stem from skewed data, flawed algorithms, or insufficient supervision in creating and deploying AI and machine learning (ML) technologies. Addressing AI bias is critical in healthcare to ensure that AI systems are reliable and equitable and do not perpetuate existing disparities in healthcare delivery. Identifying bias from the start is crucial to avoid disparities in accessing healthcare services, inaccurate diagnoses, and inappropriate treatment recommendations. These issues can have adverse effects on patients' health and well-being. In the dynamic field of machine learning and AI, it is crucial to continually train and retrain ML models as data evolves and time passes. Continuously engaging in this process enables organizations to stay alert to evolving biases and maintains the fairness and equity of healthcare systems.

Artificial Intelligence and Machine Learning

AI (Artificial Intelligence) and ML (Machine Learning) are closely related concepts that deal with developing systems and algorithms that can perform tasks that typically require human intelligence (Panch et al., 2019). Although AI and ML are similar, their approaches and scopes differ. AI is a broader concept, encompassing any system that performs tasks requiring human intelligence. ML is a subset of AI that develops algorithms that learn from data. AI and ML also address algorithmic bias in slightly different ways. In the context of algorithmic bias, AI research involves two key aspects: developing AI systems and algorithms and exploring their ethical and societal impacts (West & Allen, 2018). AI can be designed to follow ethical principles, but its capacity is limited to its training data and rules. Ultimately, ensuring ethical and socially responsible AI rests with the researchers and organizations driving its development and deployment.

In the case of AI, the simulation of human intelligence in computers or machines can include problem-solving, decision-making, natural language understanding, and more. Addressing algorithmic bias in AI involves focusing on the bias within the underlying algorithms and considering the broader ethical and social implications of AI deployment. Conversely, ML is a subset of AI that focuses on developing algorithms that enable computers to learn from and make predictions or decisions based on data. In the case of ML, algorithms improve their performance over time as they are exposed to more data (EL Bouchefry & de Souza, 2020). However, ML algorithms can also inadvertently learn biases based on inputting or calculating data (Turner-Lee et al., 2019).

AI and ML are increasingly urgent subjects to address across many industries, including healthcare. After all, these technologies have become ubiquitous. The ascent of AI and ML holds the potential for enhanced efficiency, accuracy, and, at times, impartiality. However, it is essential to acknowledge that biases in AI and ML often reflect the prejudices ingrained in our society. After all, AI is built by humans and deployed in systems and institutions (Akselrod, 2021).

Leaders must recognize the critical importance of Augmented Intelligence in ensuring that AI and ML systems undergo regular human and institutional oversight, drawing on the combined strengths of Artificial Intelligence and Human Intelligence. By fostering collaboration among diverse and inclusive groups, we can uncover and tackle biases, ultimately improving oversight and broadening the impact of these systems. With the increasing attention directed at artificial intelligence and machine learning and the biases these programs may carry, the ability of senior leaders to promote and sustain effective oversight is critical today.

In response to the ever-growing need to improve patient care and health outcomes, the adoption of AI and ML in the United States healthcare system is on the rise (Daneshjou et al., 2021). However, algorithmic bias in programs designed to assist with healthcare decision-making is associated with various challenges, some resulting in health disparities (Walsh et al., 2020). To mitigate potential health disparities and improve patient safety/care quality, multiple studies have been conducted to determine the source of the algorithmic biases and the approaches that can be adopted to address the issues (Agarwal et al., 2023; Green, 2022). Nevertheless, there is a dearth of literature regarding leadership's role in preventing algorithmic bias in the healthcare sector.

The paper's general problem is that algorithmic bias perpetuates health disparities, making it challenging to ensure health equity in the United States (Ahsen et al., 2019; Belenguer, 2022). Leadership plays a crucial role and offers significant opportunities to address algorithmic bias when collecting, analyzing, and using data in healthcare. Organizational leadership is needed to adopt ethical procedures (Kim & Cho, 2022; Walsh et al., 2020). As artificial intelligence and machine learning algorithms are increasingly integrated into various aspects of healthcare, including diagnosis, treatment recommendations, and resource allocation, ensuring these algorithms are fair, unbiased, and equitable is essential.

This paper explores healthcare leaders' role in preventing algorithmic bias when collecting, analyzing, and using data in healthcare. By incorporating these considerations, leaders can help prevent algorithmic bias in healthcare and ensure that data-driven decision-making contributes positively to patient outcomes and healthcare equity. However, to address bias, it is critical to understand and define how it can appear.

Types of Bias

In healthcare, bias can refer to various complicated and interrelated issues. Within the healthcare sector, algorithms driven by artificial intelligence and machine learning play a vital role in making crucial decisions about patient care and resource allocation; they are essential for enhancing healthcare quality and facilitating its transition (Ahmed et al., 2020). As Hoffman and Podgurski (2020) pointed out, algorithmic discrimination is not uncommon in healthcare, manifesting in race-based adjustments or identifying candidates for high-risk care management. Below are the AI and ML definitions of bias in healthcare essential to understanding the topic.

Algorithmic Bias: Algorithmic bias refers to preferences that emerge from the design and implementation of the AI algorithms themselves. Collectively, these shortcomings produce 'algorithmic bias,' which, at present, is not defined in the context of health systems (Panch T, Mattie H, Atun R., 2019). According to Yves Saint James Aquino (2023), algorithmic bias refers to the tendency of some AI systems to perform poorly for disadvantaged or marginalized groups. These biases can be unintentional and are often a result of the data used or the way the algorithm is designed. For example, an AI or ML system used to diagnose a disease may be more accurate for certain conditions but less for others, leading to unequal patient outcomes.

Data Bias: AI and ML systems in healthcare rely heavily on data for training and decision-making. Data bias occurs when skewed data is used to train AI and ML models; in other words, when the data fed into algorithms does not represent the entire population. Inaccuracies and disparities in data lead to inaccuracies in healthcare outcomes. "Bias can be manifested in (multimodal) data through sensitive features and their causal influences, or through under/over-representation of certain groups" (Ntoutsi et al., 2020). For example, if an AI system is primarily trained on data from a specific demographic group, it may need to improve when used on individuals from other demographics.

Outcome Bias: This type of bias occurs when AI and ML systems are evaluated based on specific outcomes or performance metrics that may not capture their full impact on healthcare disparities. "AI Bias is when the output of a machine-learning model can lead to discrimination against specific groups or individuals. These tend to be groups that have been historically discriminated against and marginalized based on gender, social class, sexual orientation, or race, but not in all cases" (Belenguer, 2022). This bias may occur when an AI or ML system is designed to reduce the cost of care—it may prioritize specific treatments or interventions over others, which could discriminate against specific patient groups.

Patient Bias: Patient bias occurs when irrelevant patient attributes, such as race, gender, or socioeconomic status, influence AI/ML-driven healthcare decisions. The patient experience is the "sum of all interactions, shaped by an organization's culture that influences patient perceptions across the continuum of care" (The Beryl Institute, 2023). AI or ML systems may inadvertently consider these factors when making predictions or recommendations, leading to unequal treatment for different patient populations.

Clinical Bias: Clinical bias refers to biases that arise from the historical and subjective decisions made by healthcare professionals. It refers to a set of cognitive tendencies of clinicians to make decisions based on incomplete information or subjective factors or out of force of habit (Aquino, 2023). If AI models are trained on data that reflects such biased decisions, the AI system may perpetuate the same biases, potentially reinforcing healthcare disparities.

User Bias: User bias comes into play when healthcare professionals interact with AI tools and unconsciously rely on AI recommendations without questioning their validity. This behavior often appears in the feedback loop when clinicians accept AI recommendations even if they are incorrect, leading the algorithm to relearn and perpetuate the same mistakes (Ueda et al., 2023). This can reinforce existing biases and hinder the critical thinking needed to challenge or validate AI-generated suggestions.

Algorithmic bias in healthcare encompasses systematic and unjust discrimination within algorithms used for critical functions like diagnosis, treatment recommendations, and patient management (Hoffman & Podgurski, 2020). These biases can emerge from various sources, including biased trained data, flawed algorithms, and the data collection process itself. The repercussions of algorithmic bias in healthcare are extensive and disproportionately affect marginalized communities. Hoffman and Podgurski (2020) highlighted the prevalence of "race corrections" in clinical algorithms, where developers argue that adjustments based on historical data are justified. The article *Artificial Intelligence and Discrimination in Health Care* highlighted specific instances, such as the American Heart Association's heart failure risk score assigning extra points to "nonblack" patients, potentially downplaying the risk for Black individuals. Another example involved an algorithm assessing kidney function reporting higher rates for Black patients, implying better kidney function. The Kidney Donor Risk Index suggested a higher graft failure risk for Black donors, potentially limiting their suitability. The Vaginal Birth After Cesarean algorithm predicted lower success rates for vaginal births in African American and Hispanic women, potentially leading to more surgeries. Additionally, an algorithm predicting kidney stones in emergency room patients added points for nonblack patients, potentially underestimating the likelihood of kidney stones in Black patients.

Literature Review

The primary aim of this paper is to synthesize existing research on AI and ML biases across various fields, with a particular focus on its implications for healthcare. The literature includes a great deal of research on AI and ML bias. This section will outline various subtopics, such as transparency, fostering team diversity, establishing accountability mechanisms, optimizing data management, and enhancing training protocols. The subtopics will be examined from a theoretical and ethical perspective without necessarily providing concrete, implementable solutions for each strategy.

My goal for future research is to build a comprehensive resource for healthcare leaders, policymakers, and future researchers, facilitating a better understanding of the complex issue of AI and ML bias and its impact on the healthcare sector. I will present a thorough literature review offering a detailed background in this establishing document. The following sections will explain the issue's scope and dimensions regarding ethics, theory, and strategic/practical solutions. In the analysis, I will identify and elaborate on major emerging themes from the literature review and then include policy recommendations as a conclusion.

Leadership Strategies to Reduce Bias

Various authors offer unique perspectives on the responsibilities and best practices for leaders working to reduce bias. According to Mello et al. (2022), leaders of organizations are responsible for mitigating algorithmic biases within their internal decision-making models. Cognitive biases and barriers to rational and objective decision-making lead to systemic patterns of deviation from rationality when using AI-powered decision-making tools and systems (Mello et al., 2022). In detail, leaders should identify any bias and understand where it occurs, train employees to address it and collaborate with other mitigants to improve the likelihood of the AI-powered tools' success.

Additionally, according to Boniface et al. (2022), executive leaders should embrace and implement secure, rights-respecting approaches endorsed by the entire community and other stakeholders to initiate trust in providing sensitive health data to clinicians. Leaders must bridge the gap between data and trust among patients, healthcare service providers, and other users for higher levels of data stewardship, citizen participation, and acceptance. For example, significant sources of patient-related data can be acquired through wearables, environmental screening, social media, and other healthcare platforms (Boniface et al., 2022). However, the quality of care provided in different clinical settings entirely depends on the quality of patient data, which the owners are only willing to provide if they are ensured data security and privacy. What about those who choose not to enter specific data?

Fejerskov (2021) also acknowledges that algorithmic bias is a significant vulnerability of automated systems in the healthcare sector; it perpetuates and exacerbates existing discrimination based on ethnicity, gender, and race. Unfairness in diagnoses and prognoses aided by AI-powered decision-making tools and systems affects marginalized groups negatively, leading to unhealthy outcomes (Fejerskov, 2021).

Vayena et al. (2018) acknowledge that ML algorithms used in medicine and clinical settings are coupled with several ethical challenges. Therefore, they recommend that executive leaders commit to addressing these issues with the help of employees, developers, and other healthcare stakeholders. The Vayena et al. (2018) study suggests that leaders should take concrete steps to address any possible ethical challenges of privacy, trust, and accountability associated with machine learning algorithmic bias. For example, during training, data sourcing, and collection, leaders should ensure that developers and other stakeholders adhere to data protection and privacy requirements to safeguard patients' sensitive and confidential data. Data protection has been inconsistent or not addressed, hindering digital health adoption and reducing research opportunities due to damage to an organization's reputation or even legal and regulatory consequences. According to Vayena et al. (2018), computer science is based on the principle of "Garbage In, Garbage Out" (GIGO). Therefore, the quality of the data used in training and validating different AI algorithms in healthcare significantly predicts the quality of outcomes the AI tools will have on the entire clinical setting.

According to Köchling et al. (2021), biased input data causes biased algorithms, which results in unfairness and disparate treatment for different groups. Data containing explicit and implicit stereotypes, prejudices, and human judgments lead to inaccurate findings and conclusions. In contrast, data curated without explicit or implicit stereotypes, prejudices, or human judgments leads to accurate findings and sound conditions. Leaders should be aware that the data used to train algorithms is based on past data, which might contain biased decisions, perpetuating racial disparities and inequalities (Agarwal et al., 2023). Algorithms are often crafted using historical or preexisting data as their bedrock. This practice of training algorithms with past data is an important, lasting tradition. As a result, leaders must remain vigilant, recognizing that the data used for algorithmic training might harbor biases. These biases can perpetuate ongoing racial disparities and inequalities if left unaddressed.

It is crucial to incorporate the understanding that structural racism is at the core of racial health disparities, and research underscores the pivotal role of leaders in tackling challenges related to bias in artificial intelligence (AI) and machine learning (ML) within the healthcare sector (Cerdeña et al., 2020). Addressing this complex issue demands a comprehensive approach encompassing several vital dimensions. The following recommendations outline strategies for mitigating AI and ML bias by emphasizing transparency, fostering diversity within teams, establishing accountability mechanisms, optimizing data management, and enhancing training protocols.

Transparency

Leaders must implement practices ensuring integrity and transparency throughout data collection, analysis, and patient communications. Ensuring the quality, efficiency, and continuity of data throughout its life cycle, data integrity in the healthcare sector encompasses safeguarding patient information, health reports, diagnostic records, laboratory test reports, and other relevant records (Zarour et al., 2021). Bernal and Mazo (2022) propose that improving AI transparency is a significant method of addressing trust issues within the clinical setting and reducing harm from unfair machine learning algorithms. “Transparency” in the Bernal and Mazo (2022) study is an umbrella term for privacy, security, intellectual property, equity, and interpretability. According to the study, executive leaders should focus on improving the transparency of all AI systems used in clinical settings to improve their acceptance by both practitioners and patients. Additionally, emphasizing transparency underscores the significance of fostering open communication and collaboration with various stakeholders, including patients, regulators, and the general public.

Executive leaders in healthcare settings can also bridge the data-trust gap by upholding a social license. According to Boniface et al. (2022), a social license refers to the permission granted to researchers by the community to collect, use, and share healthcare data that is useful in research. This license is offered in response to trustworthiness (i.e., benevolence, integrity, and accountability) on the part of the researchers and adherence to the values and expectations of communities and the public. In simpler terms, social license depends on the communal perception that whatever is done is beneficial, ethical, and legally acceptable. Accordingly, leaders should show that they are committed to protecting the privacy of the individuals whose data is being used in algorithmic development and training to improve the quality of health care provided.

Ethical communication and patient interactions with patients achieve transparency and the development of a social license. According to Vayena et al. (2018), developers and medical practitioners must disclose basic details about specific tools and programs to patients to satisfy transparency needs, a core element of medical ethics. Leaders should follow the Health Insurance Portability and Accountability Act (HIPAA), the Electronic Communications Privacy Act (ECPA), and the Children's Online Privacy Protection Act Of 1998 (COPPA) guidelines to protect the privacy of owners' data while also implementing fundamental data management and population health capabilities such as de-identification, identity masking, population stratification, and generalization (Lobban, 2022). These practices ensure that data is preserved while safeguarding individual privacy. Consequently, more patients will be willing to share their data to improve algorithm training and achieve fairer healthcare outcomes (Hryciw et al., 2023).

Strategic steps to increase transparency occur at each stage of the machine-learning process. When choosing or designing an AI system, leaders should ensure the AI system's interpretability by ensuring data scientists can adequately discuss the algorithms’ decision-making process (Bernal & Mazo, 2022). When this occurs, end users of the AI systems will amass more confidence in using the AI tools and systems within clinical settings. Moreover, developers must have evidence (Badal et al., 2023) that supports the safety and efficacy of any AI-powered devices within clinical settings. For instance, the FUTURE-AI (Fairness et al.) medical algorithm checklist is an example of a potential clear, thorough, and systematic way that translates high-level AI principles and models into practical computational guidance (Badal et al., 2023). Medical AI-powered tool deployment must also satisfy transparency standards. Köchling et al. (2021) hold that executive leaders should regularly implement proactive auditing methods to verify and audit all algorithmic decision processes. These automated auditing tools can effectively assess the context-specific fairness of algorithmic decision-making tools and machine learning systems (Mökander et al., 2021).

Finally, transparency within and among the medical community is a significant bias reduction tool. Executive leaders must proactively and rationally engage in cognitive bias prevention and impact mitigation, for example, cross-checking conclusions with professionals outside their development and implementation teams to gain a second and unbiased perspective of any questionable assumptions (Mello et al., 2022).

Fostering Diversity

Leaders can address preexisting bias by ensuring diversity and inclusivity in development teams since such bias mainly originates from individuals with significant input into developing AI or ML technology (Pamch et al., 2019). Jackson (2021) proposes that executive leaders and other stakeholders should focus on increasing diversity in the technology industry to minimize and address human bias from the people creating the algorithms. When people with diverse backgrounds are involved in collecting training data, they may be able to moderate the data and eliminate possible human bias effectively. Comprehensive and reliable data sets for training and validation are imperative in making the algorithm robust and reducing biases. Prioritizing diversity and inclusion within the project development and deployment team supports objectivity, improves critical thinking, prevents unconscious or implicit cognitive biases, and enhances project outcomes (Panch et al., 2019).

From an ethical perspective, diverse teams will effectively address implicit and unconscious biases by representing the interests and grievances of the groups they represent. According to Simon et al. (2020), technical and emergent bias can be effectively mitigated through initiating bias monitoring and correcting systems and models capable of identifying biases as soon as they occur and correcting them amicably. According to Fejerskov (2021), ensuring an inclusive and diverse development team reduces the risk of overestimating health risks and issues associated with a specific group. Similarly, such groups are well-positioned to collect quality and accurate data, reflecting the health aspects of a broad spectrum across ethnicities, races, genders, and socio-economic classes (Fejerskov, 2021).

When strategically building their teams to foster diversity, leaders have several aspects to consider. First, leaders should insist that any algorithm being developed has a basis in adequate medical knowledge—requiring those working on the algorithm to possess a solid understanding of medical concepts and principles to ensure the algorithms are accurate, effective, and safe. Strategically, according to Bertl et al. (2023), during the development of any medical algorithm, leaders must engage experts from different professional fields and allow them to bring their expertise to different subject matters. By doing this, the experts investigate all essential features of the machine learning algorithm being developed from an organizational, social, and technical perspective (Bertl et al., 2023). These strategies reduce misrepresentation and bias associated with inadequate datasets.

Accountability

Executive leaders should not rely solely on the information provided by algorithms and implement their decisions without significant human control (Köchling et al., 2021). Auditing is an inherently dynamic process, emphasizing ongoing proactive assessments that necessitate human oversight; fostering collaboration with diverse auditing teams and incorporating external perspectives alongside internal employees enhances independence and provides a broader external viewpoint (Landers & Behrend, 2023). To hold their teams (and themselves) accountable, leaders should have a solid understanding of bias and the process to address it. Mello presents a five-step approach for healthcare executives working to address cognitive biases. First, identify and define the problem of biases and address any concerns and questions. Second, gather and evaluate information regarding algorithmic biases and their impact within healthcare settings. Third, analyze the information gathered. Fourth, when it comes to making decisions, one thing that significantly enhances accountability is the ability to explain what led to the decision being made and what factors influenced the outcome. This principle of explainability is not only crucial in the realm of artificial intelligence but also in broader decision-making contexts. When decisions are made based on clear, transparent explanations, it fosters trust, understanding, and ultimately accountability. Incorporating this principle into decision-making processes helps executives make well-founded conclusions based on practical and accurate data analysis findings. Fifth and lastly, there should be articulation and documentation of the rationale for the decision for future reference and guidance. Following these steps of inquiry and documentation will allow future leaders to deconstruct their processes.

Kallu et al. (2022) point out that algorithmic biases can occur in many ways, even if the data used for training is fundamentally sensitive and adequately scrutinized for inconsistencies. According to their study, leaders should find methods of assessing algorithmic fairness in decisions, identify any cases of inconsistencies, and initiate procedures to solve the problem before it is too late. Similarly, Jackson (2021) articulates that there is a higher probability of bias when algorithms are inaccurately produced by overrepresenting a particular dataset or underrepresenting data associated with specific groups. For example, algorithms are used in contemporary medicine to identify which patients need special treatment the most (Jackson, 2021). However, many AI-powered tools and systems using biased algorithms have racially discriminated against and excluded African Americans instead of being race-neutral as they anticipated. According to Bertl et al. (2023), despite the notion that the acquisition of AI in the healthcare sector is intended to improve accuracy, efficiency, and patient-centered services, algorithmic bias and uncertainty lead to poor medical care.

According to Mello et al. (2022), early bias identification and remediation minimize the impact of the bias within an AI system's project life cycle, including subsequent updates and refreshes. Ultimately, leaders should deploy algorithmic bias monitoring, which involves continuously analyzing the outputs of algorithms and ML models to detect and mitigate potential biases that may unfairly impact individuals or groups based on sensitive attributes. Leaders should also look to deploy identification models that involve proactively tracking and verifying the identity of individuals or entities through various methods, such as biometrics, authentication processes, or data validation, to ensure secure and accurate access to resources or services. Jackson (2021) proposes that leaders strategically initiate periodic internal audits to monitor, identify, and adjust algorithms with discriminatory and unfair outcomes. Auditors must conduct proactive monitoring and verification to promptly identify and address any bias in algorithms, ensuring that potential issues are mitigated or resolved before they escalate into significant problems. Moreover, leaders should ensure they initiate and fund AI bias awareness processes and models aimed at bias identification and mitigation to avoid exacerbating current inequities associated with unfair algorithmic decisions effectively. In other words, executives should commit to ensuring internal accountability to enhance fairness and remediate algorithmic biases.

Data Management

According to a study on bias and fairness in machine learning conducted by Mehrabi et al. (2021), fairness is the absence of any favoritism or prejudice towards an individual or a group based on acquired or existing characteristics. AI algorithmic bias compromises fairness, a significant issue when developing and applying ML in different critical sectors. Mehrabi et al. (2021) propose tactical methods for fair machine learning, divided into pre-processing, in-processing, and post-processing. In detail, pre-processing helps change data to do away with or remove underlying discrimination in training data. In-processing modifies learning algorithms to remove bias and unfairness in the model training process by changing objective functions or imposing constraints (Mehrabi et al., 2021). The ongoing development of machine learning models necessitates an adaptable strategy for retraining and post-processing. Regarding retraining/deep learning, it is crucial to base decisions on data and consider multiple factors (Panch et al., 2019). Simultaneously, post-processing offers a valuable means of enhancing model outcomes and upholding machine learning applications' effectiveness and ethical integrity. Managing the interaction between retraining and post-processing is vital to upholding model performance and reliability within ever-changing environments. As an illustration, this might entail classifying algorithms as "black boxes" when neither system, tool training data, or algorithm functions can be modified (Panch et al., 2019, p. 2). According to Čartolovni et al. (2022), black-box algorithms (opaque processing systems that do not share their data analysis processes) significantly impact the fiduciary relationships between healthcare practitioners and patients. In other words, black-box algorithms work by taking in some ingredients (data) and following hidden steps to produce a final dish (output or decision), but they do not reveal how they cooked it. You can judge the meal (results) but cannot see how it was prepared. This "secret recipe" system can lead to mistrust and doubt.

The ethical ramifications of data collection are immense. A study by Agarwal et al. (2023) posits that algorithms trained using data that feature existing biases in the diagnosis and prognosis of marginalized populations pose a danger of exacerbating existing discrimination and adverse health outcomes. To address algorithmic bias, leaders and other stakeholders should focus on using quality data and being proactive about the data input into an AI system. For example, data used in AI-powered healthcare decision-making tools should be accurate, representative, and free from systematic bias. Quality data includes diverse patient samples, accurately represents various demographics, and thus ensures equitable and fair predictions. On the other hand, non-quality data may be biased, unrepresentative, and lack accuracy; non-quality data might have incomplete patient records or be skewed towards specific groups, leading to biased AI models that could perpetuate health disparities. Suppose the datasets used to train a specific algorithm do not reflect the true epidemiology of a specific group, or an algorithm is trained using information with only a few members (small sample size) of a given demographic. In that case, there is a higher probability of having disparate treatment of certain groups (Vayena et al., 2018). Such algorithms can exacerbate existing health disparities. Therefore, policies should be implemented to ensure that data used for training adequately and accurately represent the source population during algorithm development.

Given that many leaders prioritize cost reduction and efficiency enhancement, which drives the increased adoption of AI and ML in healthcare, it becomes imperative for leaders to implement practical strategies for enhancing data collection and management to optimize these systems. Leaders should analyze and control the training datasets used for any algorithms to identify and address any instances of cognitive and explicit biases. Similarly, according to Agarwal et al. (2023), a practical way to mitigate biased data is through a representative data collection mechanism where the development team is inclusive and diverse, improving fairness without sacrificing accuracy. They argue that leaders can solve the algorithmic bias problem by initiating inclusive algorithm development and testing processes, initiating bias checking, correction, and mitigation programs, and offering feedback on the impacts of the bias and the way forward. This approach addresses algorithmic bias as identified, focusing on the problem rather than support models (Agarwal et al., 2023). These practical decisions entail collecting and analyzing additional data from underrepresented groups, such as African American patients of low socio-economic status, to be used as algorithm training datasets (Agarwal et al., 2023). When deployed, algorithms with a more diverse dataset offer more fair diagnoses and prognoses, thus improving healthcare outcomes among marginalized groups (Agarwal et al., 2023).

Training Protocols

According to Simon et al. (2020), leaders with a good understanding of computer system bias have an adequate awareness of biases, allowing for easier identification of potential harms, means to avoid them during the design phase, or ways to correct them when the systems are in use (Simon et al., 2020). To achieve these aims, leaders must implement and teach an objective anti-bias decision-making methodology (Mello et al., 2022). However, a lack of training affects leaders, all employees, and even care recipients.

Alongside the bias and unfairness in IT-powered decision-making tools in clinical settings, low IT literacy, inadequate training and support, and insufficient funding have discouraged adopting and using electronic health (e-health) services (Bertl et al., 2023). Scheduling development and training workshops to investigate underlying assumptions thoroughly is a strategic step toward addressing cognitive biases in healthcare.

Strategically, executive leaders in the healthcare sector must implement training across the entire workforce to initiate bias awareness and propose methods to mitigate the impact of biases as early in the process as possible. According to Luaces et al. (2022), cognitive bias workshops and training offer insights into identifying unconscious bias, its impacts, and potential solutions. According to the study, such training workshops improve general awareness about biases, healthy interpersonal communication, and relationships between physicians and patients, especially individuals or communities facing marginalization or discrimination. These training workshops also provide providers with information that can be applied intentionally to improve services and decisions facilitated by AI-powered tools (Luaces et al., 2022). The bottom line is that training workshops can and should be engaging and relevant to contribute towards increased knowledge.

An intentional focus on transparency, diversity, accountability, data management, and training protocols using the strategic steps above allows for complete life-cycle bias reduction. Thomasian et al. (2021) articulate that AI algorithmic bias mitigation should not end at the product development phase but should continue within the entire product life-cycle. AI-powered tools' and systems' life cycles extend through development, validation, implementation, maintenance, updates, and the post-maintenance period (Thomasian et al., 2021). Notably, bias can enter any of these processes; thus, a well-positioned bias mitigation framework must cover each life-cycle phase. These solutions are also interrelated. For example, according to Thomasian et al. (2021), leaders should ensure quality and adequate data for algorithmic training to address bias associated with underrepresenting or overrepresenting some health aspects for different groups. Having diverse development teams who openly scrutinize and eliminate biased cases can help to achieve quality and adequate data.

Analysis

Diversity and inclusion in algorithm development teams is an overarching theme in addressing algorithmic bias. From a tactical perspective, data collection should involve ensuring diverse and representative data sources, while data analysis should focus on using fairness-aware algorithms and conducting regular bias monitoring. From a strategic standpoint, leaders must establish ethical data governance policies, foster collaboration, and promote awareness within diverse teams to prevent algorithmic bias in practical decision-making. According to many articles, algorithmic bias exacerbates preexisting bias consciously or unconsciously due to the overrepresentation or underrepresentation of a specific group and its interests within the development team, leading to adverse healthcare outcomes (Simon et al., 2020). Ensuring diversity and the inclusion of underrepresented or marginalized groups such as racial minorities, females, and people of low-income socio-economic status can reduce algorithmic bias during the development phase. The policy suggestions in the paper align with multiple theories of inclusivity and diversity, as discussed below.

First, the theory of generative interactions suggests that inclusion can be increased by promoting inclusive practices, culture, and behaviors that stimulate new ways of managerial strategy around inclusion (Bernstein et al., 2020; Beaudouin-Lafo, 2021). The theory postulates that inclusion and diversity within workplaces have benefits associated with enhanced social justice, equity, personal development, and entire organizational performance. On the other hand, Bernstein et al. (2020) note that organizations that lack inclusion and diversity suffer in many ways. A diverse group that collaboratively challenges systemic and structural inequities will be more successful by elevating practices of diverse representation. Inclusion and diversity facilitate generative interactions within teams, creating social connections and the profound understanding required to establish equality at the organizational level (Beaudouin-Lafo, 2021). Within generative interactions, there is a capacity to support or challenge the integral assumptions of contemporary subject matter and trigger reconsideration of ideas taken for granted (Bernstein et al., 2020). In healthcare algorithms, generative interactions enhance the collective scrutiny of training data and other aspects of AI and ML-powered programs to eliminate pre-existing implicit or explicit bias. Generative interactions within a team mitigate self-stigma, stereotypes, and prejudices and uphold collective and collaborative thinking and problem-solving skills essential in identifying and solving algorithmic bias (Turner-Lee et al., 2019).

The stakeholder theory of inclusivity focuses on those working on the algorithm, who should understand medical concepts and principles of inclusive decision-making and organizational performance (Bernstein et al., 2020). Based on Bernstein et al. (2020) and Beaudouin-Lafo's (2021) findings, executive leaders are tasked with generating more value for stakeholders without creating tradeoffs while ensuring that stakeholders' interests are similarly aligned. Representing people from different cultural affiliations within a single social system—and the feeling that every individual is an integral part of the system—reduces systematic disparities through collaboration and coordination among individuals representing diverse groups. Accordingly, the stakeholder theory advocates for fair, honest, and equal treatment of all stakeholders to achieve equitable outcomes and ethical management of organizations (Bernstein et al., 2020).

Finally, organizational justice theory conceptualizes fairness in algorithmic ecosystems (Colquitt, 2012; Kordzadeh & Ghasemaghaei, 2022). For this theory, justice refers to the perception of an organization's top management decisions as consistent with equality, respect, consistency, and trustworthiness (Kordzadeh & Ghasemaghaei, 2022). Procedural and distributive justice are two primary components of organizational justice, which depend on the strategic initiatives proposed and advocated by the executive leadership. According to Kordzadeh and Ghasemaghaei (2022), distributive justice occurs when decision outcomes within an organization align with existing equality and equity norms and the need-based allocation of harms and benefits. Conversely, procedural justice involves making decisions that uphold equitable processes, including accuracy, consistency, ethics, and equal representation (Lambert, 2020; Kordzadeh & Ghasemaghaei, 2022). In algorithmic bias, procedural justice and fairness can be achieved through leaders, including system interpretability and transparency in the decision-making process (Kordzadeh & Ghasemaghaei, 2022). Similarly, distributive justice can be achieved by engaging diverse groups, including marginalized ones.

Another overarching theme in the literature review is the role of leadership in the algorithmic bias mitigation processes. According to Sims Jr. et al. (2009), based on the situational leadership theory, the best leadership style for an organization depends on the immediate situation created by the present internal and external environment. From the comprehensive review above, it is clear that many healthcare stakeholders, such as patients, are reluctant to provide their sensitive data to developers due to privacy and data security issues.

Based on the situational leadership theory, a good leader should understand how to lead others to be good leaders through effective behavior (Sims Jr. et al., 2009; Thompson & Glasø, 2015). Leaders should empower employees and other stakeholders by transparently outlining the available tools to enhance privacy and data security when tensions and uncertainties around data privacy and security are present. Leadership philosophies tied to influencing others toward a common end are effective in achieving fair and unbiased algorithms. According to Sims Jr. et al. (2009), transformational leadership through inspiration and motivation is the best mechanism for empowering stakeholders to share their data, which is essential in algorithm training to solve algorithmic bias. Ideally, the motivation induced by executive leaders will prompt support of periodic algorithmic bias monitoring and auditing models. When the top leaders in a company inspire and encourage their employees to understand the theory behind AI and ML tools and the inherent ethical issues tied to bias within those tools, employees will regularly check and repair any biases in the algorithms. Leaders must cultivate a dedicated team that monitors algorithms to ensure fairness. This involves nurturing a workforce that is both motivated and deeply committed, instilling a sense of ownership among them. The leader fosters such an environment and ignites the passion to achieve this goal effectively.

Similarly, an empowering leadership style is a practical method to ensure diversity and inclusivity in algorithm development teams as the team becomes more experienced. Boal and Hooijberg (2000) contend that strategic leaders can foster the growth of essential skills and capabilities in their colleagues, thereby generating ethical value and maintaining organizational performance. Strategic leadership entails creating and maintaining absorptive capacity (the ability to learn and apply new knowledge toward an unknown end) and adaptive capacity (the ability to change with the prevailing environment) (Zhang et al., 2010). When these capacities are reinforced with managerial wisdom, organizations are better equipped to weather new problems and situations (Castillo & Trinh, 2019).

Based on the upper-echelon theory, Boal and Hooijberg (2000) posit that organizational relationships and outcomes immediately reflect executive leaders' cognition, strategic choices, and values. Therefore, to confront algorithmic bias within the healthcare setting, leaders must effectively invest their resources in learning new information about AI and ML and sharing their knowledge with their juniors to enhance positive organizational outcomes, for example, through training workshops. These workshops are necessary for imparting knowledge to the workforce on the possible causes of algorithmic bias, its impacts, and potential bias mitigation mechanisms and strategies (Mello et al., 2022). Educational seminars and training workshops for healthcare stakeholders on ML– and AI-powered tools and systems development, deployment, and management align with social learning theories.

Bandura's social learning theory explains that human behavior is shaped through cognitive, environmental, and behavioral interactions within an immediate social context. Therefore, learning can occur through observing one's immediate environment and modeling (Chowdhury, 2006; McLeod, 2011; Rumjaun & Narod, 2020). Within the healthcare setting, executive leaders are supposed to be role models from whom subordinate employees can learn and emulate. According to Chowdhury (2006), employees are likelier to copy or learn from their superiors than their peers. Therefore, initiating training and development seminars and workshops involving leadership is a better approach for imparting positive behavior among the workforce and solving existing algorithmic bias.

Finally, the literature review above illuminates the effects of the immediate organizational environment in identifying and solving algorithmic bias within a clinical setting. Periodic and standby audits on AI-powered tools and systems to monitor biases are essential for prompting timely and effective mitigation of impacts associated with algorithmic bias. The idea that bias monitoring should occur throughout the entire lifecycle of ML– and AI-powered programs aligns with the stimulus-organism-response theory (Kordzadeh & Ghasemaghaei, 2022). This theory postulates that environmental stimuli trigger one's internal states, influencing a behavioral response (Nagoya et al., 2021; Kordzadeh & Ghasemaghaei, 2022). The stimulus-organism-response theory can be used to explain executive leaders' behavior in dealing with algorithmic biases within the clinical environment. In this context, the *stimulus* refers to identified instances of algorithmic bias within an AI-powered tool or system, and the *organism* refers to the anticipated fairness. The *response* is the behavioral reactions of leaders and other stakeholders in solving the existing bias amicably (Kordzadeh & Ghasemaghaei, 2022). The findings of this paper suggest several effective behavioral reactions by leaders: building inclusivity and diversity in algorithm development teams, promoting readiness to address any possible bias throughout a system's lifestyle, and establishing periodic and comprehensive bias detection and mitigation processes.

Ethical Implications of Algorithmic Bias

According to Martinez-Martin et al. (2020), the continuous use, storage, and analysis of patient data using AI-powered tools vulnerable to algorithmic bias raises ethical issues associated with data protection, privacy, confidentiality, informed consent, and fairness. Existing ethical and regulatory frameworks may fail to address these issues. Luxton (2020) explains that algorithmic bias occurs when a program makes systematic errors that lead to unfair outcomes, including favoring one group. Significant sources of algorithmic bias include missing data, measurement errors, misclassifications, and small sample sizes, resulting in underestimation and inaccurate predictions for subgroups of patients. Therefore, algorithmic bias in healthcare ML and AI-powered decision-making tools, such as chatbots and other conversational agents, may lead to privacy issues if left unsolved. Some AI-powered tools, like chatbots, may collect comprehensive amounts of private and sensitive patient information (Luxton, 2020). However, conversational agents can and should inform end users of existing limits to privacy and available protections for private and confidential patient data. Privacy interests in the US are protected by constitutional law, professional ethics, state statutes and regulations, and cultural norms (Martinez-Martin et al., 2020).

Ethically, privacy encompasses various obligations to protect an individual or their data from unwarranted intrusions (Martinez-Martin et al., 2020). Voice, facial attributes, gait, heart rate, biometric information such as fingerprints, and data that can reveal one's IP address and physical location are considered private in clinical settings. This information needs to be protected from unwanted and unauthorized disclosure. However, using ML– and AI-powered tools such as ambient sensors and monitoring devices may undermine personal informational privacy if their data is leaked without the owner's consent.

Apart from informational privacy, algorithmic bias also compromises patients' decisional privacy. As Martinez-Martin et al. (2020) postulate, decisional privacy is the right offered to individuals to determine the nature and type of health care they should receive without interference from the government or other stakeholders. However, algorithm bias in many ML– and AI-powered decision-making tools will significantly compromise decisional privacy, as they collect personal data and analyze it based on data used in their training to project and recommend the type of care best for a specific patient.

Another practical ethical issue surrounding the application of AI-powered tools is the issue of inequitable access. According to Luxton (2020), accessibility of healthcare technologies depends on technological infrastructure investment in various communities and regions. Similarly, it depends on technological literacy among communities of distinct socio-economic conditions. Therefore, people who lack adequate access to healthcare technologies that collect healthcare data, such as conversational agents (medical chatbots or healthcare chatbots), end up experiencing healthcare disparities compared to others within the population (Luxton, 2020).

It is important to note that the accuracy of machine learning algorithms depends entirely on the correctness and quality of their training and validation datasets (Martinez-Martin et al., 2020). Therefore, the absence of data associated with a particular population sector may lead to systemic algorithmic bias. If an AI-powered decision tool's algorithm is trained using incorrect or biased information, there will be biased healthcare outcomes (Martinez-Martin et al., 2020). Such tools lack sensitivity to and awareness of cultural and other socio-economic differences in behaviors, leading to disparate treatment among patients, especially those from racial minorities.

Other ethical issues expected to arise when algorithmic bias is left unsolved in clinical settings include liability, data fairness, transparency, and accountability (Vayena et al., 2018). The issue of liability arises when a biased computer program causes harm to a patient in healthcare. Determining who should be responsible—the program developer, the healthcare provider using it, or both—is complex and unclear. Legal and ethical challenges arise due to the absence of well-defined rules in such cases. There are often insufficient laws to assign blame in medicine when biased algorithms lead to harm, causing uncertainty and unfairness for patients. Accountability in AI, particularly in healthcare, is problematic because the decision-making process of many AI models is opaque, making it difficult to understand how and why certain decisions are reached. When these systems introduce or exacerbate bias, holding anyone accountable for the outcomes becomes challenging. Additionally, it is often unclear who should address bias—developers, healthcare providers, or regulators—leading to a lack of accountability. This lack of accountability can result in ethical concerns regarding fairness, liability, transparency, and data reliability in clinical settings. All these issues scare away healthcare stakeholders, including doctors, trainees, and patients, from accepting the implementation of AI-powered decision-making tools within the clinical setting (Hryciw et al., 2023).

Social Effects of Algorithmic Bias on Society

Discriminatory algorithmic outcomes that exhibit unfair or prejudiced behavior toward certain groups of people can have significant social effects on society, leading to various challenges and negative consequences. According to Čartolovni et al. (2022), social issues related to algorithm bias in healthcare should not be overlooked. Medicine or healthcare is a significant social practice entirely affected by various social determinants of health. For instance, Čartolovni et al. found that existing social disparities and inequalities can be significantly replicated in healthcare AI-powered tools through AI algorithm bias, leading to further inequality, potential discrimination, and more adverse healthcare outcomes among already marginalized groups (Čartolovni et al., 2022). According to Lysaght et al. (2019), healthcare practitioners use AI-assisted decision-making tools to carry out diagnoses and predict treatment outcomes on patients based on available relevant data, such as social and medical history, diagnostic tests, socio-demographics, and genome sequences contained in patients' electronic health records. Walsh et al. (2020) posit that algorithmic bias is a significant cause of self and public social stigma, contributing to adverse healthcare outcomes. Self-stigma then leads to self-discriminating and stereotyping behavior among patients.

Additionally, public stigma compromises proper healthcare providence and can have far-reaching consequences. It can negatively impact the quality of care given to stigmatized individuals, result in coercive treatment approaches, and even contribute to punitive policies against those facing stigma. For instance, punitive policies may be imposed on stigmatized groups instead of fostering understanding and support. Likewise, healthcare providers influenced by public stigma may resort to forceful treatment methods, neglecting patient-centered care rooted in informed consent. Coercive treatment and punitive policy-making in healthcare broadly cover enforcing medical interventions and policies on individuals without their full consent, often through force, threats, or corrective measures. Coercive treatment involves compelling individuals to undergo medical procedures against their will or without informed consent (Hem et al., 2018). Forming punitive policies involves establishing regulations and laws that impose significant consequences for failure to adhere to healthcare guidelines. This may encompass measures such as compulsory medication, involuntary hospitalization, the reinforcement of stigma and discrimination against health conditions, and the criminalization of specific health issues.

Another significant social issue that emerges due to failures in addressing algorithm bias in AI-powered healthcare tools is adverse patient-physician relationships. According to Čartolovni et al. (2022), using AI-powered decision-making tools such as black-box algorithms significantly impacts the fiduciary relationships between healthcare practitioners and patients. These tools may undermine trust in patient-physician relationships in contemporary society. Similarly, a Martinez-Martin et al. (2020) study found that AI-powered tools with algorithmic bias can potentially be misused, damaging clinical relationships between physicians and patients. For example, medical professionals might rely on inaccurate AI outputs, leading to potential misdiagnoses or inappropriate treatments of patients from marginalized communities.

Other social issues associated with AI-powered tools include public trust and acceptability. Although the goal of implementing AI in healthcare is to improve diagnostic and prognostic efficiency, algorithm biases can significantly undermine its acceptance (Čartolovni et al., 2022). According to Čartolovni et al. (2022), negative opinions and prejudices about AI superiority and singularity exist among healthcare practitioners, significantly undermining the AI-powered tool's acceptance within healthcare settings. Some healthcare practitioners have opposing views and biases about AI's and ML's potential in healthcare, particularly regarding concerns about AI surpassing human capabilities and operating autonomously (Khan. et al., 2023). These biases exist within the healthcare community and hinder the acceptance of AI and ML in healthcare settings. Ultimately, if left unaddressed, these biases could result in algorithmic biases in healthcare AI systems. This, in turn, may erode trust in patient-physician relationships and impact both current and future medical practices (Panch et al., 2019). If left unchecked, algorithm biases can result in shallow and deficient trust in patient-physician relationships in contemporary and future medical practice (Hryciw et al., 2023).

Policy Recommendations

Policy creation to address algorithmic bias can be grouped into three broad categories. According to Vayena et al. (2018), these categories include data sources for algorithm training, machine learning algorithm development, and deployment of the trained and approved algorithms in the clinical setting. In detail, training data sources must adhere to the data protection and privacy requirements of existing ethical and legal frameworks to address algorithmic bias. Healthcare algorithm development should be committed to fairness, and its deployment must satisfy available transparency standards (Vayena et al., 2018). According to Badal et al. (2023), AI developers should follow best practices during the AI tools development phase. During development, AI teams should commit to achieving fairness in how algorithms analyze data sourced from patients from various demographics and socio-economic conditions.

To adhere to specified data protection and privacy requirements, developers must pay close attention to the ethical and legal restrictions of every aspect of the data collection, analysis, and processing stages. Data collection must be authorized through owners' informed consent to use or reuse their data in algorithm training. Regulators in different jurisdictions, including states and local governments, must adhere to available frameworks such as the US Health Insurance Portability and Accountability Act (HIPAA). HIPAA commits to ensuring the privacy and accountability of data sourced from patient records, either conventional or contemporary electronic health records. Medical regulatory bodies such as the World Health Organization and the U.S. Food and Drug Administration must create and embrace review boards, ethics review committees, and medical technologies assessment organizations to check AI-powered devices' compliance with set policies and regulations (Vayena et al., 2018). Similarly, developers should commit to ensuring timely updates of various AI-powered tools and devices within healthcare settings based on technological advancements or pitfalls identified after informed assessments. Such medical bodies can also invest in funding the development of more representative datasets used for training and validation. According to Vayena et al. (2018), developers must also create standard procedures for effective post-market monitoring mechanisms to document the evolution of specific AI software transparently.

Summary and Conclusion

Healthcare leadership's role in mitigating and addressing algorithmic bias is essential. When leaders know their primary role in solving algorithmic biases within healthcare settings, all phases of the development of healthcare algorithms can be equipped with the policies and people needed to prevent bias. This research outlines the theoretical backgrounds of multiple sources of algorithmic bias within healthcare AI-powered tools and systems, the ethical impacts of that bias, and practical bias mitigation mechanisms within the reach of healthcare leadership.

According to the literature review, improving ML and AI-powered systems' interpretability and transparency effectively addresses trust issues surrounding unfair machine learning algorithms in healthcare settings. Transparency increases the willingness of healthcare practitioners to use ML and AI decision-making systems. Furthermore, in scenarios of increased transparency, patients are willing to share their data, including the most critical and confidential, to address algorithmic biases for fairer healthcare outcomes. Ultimately, increased interpretability grants healthcare officials a social license to collect, use, and share helpful healthcare data, reflecting the exact epidemiology of a specific group and healthcare stakeholders.

Leaders should also clearly designate specific employees to be responsible and accountable for applying algorithmic decision-making tools and systems once they have been designed. These employees should have a similar diversity in background, education, and gender as the development team. Moreover, healthcare leaders must ensure a diverse and inclusive team that adequately represents all critical groups served by the algorithm being developed. Fair representation would ensure that training data reflect a specific group's true epidemiology, an adequate sample size of each particular demographic, and the continuous cross-checking of facts to eliminate possible unconscious biases.

Likewise, the algorithm's development team can effectively identify and correct any instances of biased training data and eliminate inconsistencies when given the right tools. To achieve internal accountability, leaders should deploy algorithmic bias monitoring, identification models, and internal system auditors to facilitate immediate identification, reporting, and mitigation of algorithmic biases. Leaders should also display trustworthiness through a commitment to ensure patient accountability in securing the data provided by individuals during algorithmic development to uphold confidentiality, security, and fairness.

Biased algorithm training data is among the most common sources of algorithmic bias in healthcare settings. Therefore, quality and trustworthy training data is the only solution to algorithmic biases from biased training datasets. The problem can be significantly reduced by ensuring patient data privacy and security to earn the trust of data providers, who, in turn, will provide quality and correct data.

Finally, workforce training on algorithmic bias sources, impacts, and mitigation mechanisms strategically addresses algorithmic biases. General awareness of algorithmic biases supports bias identification and mitigation while enhancing critical thinking and objectivity—eliminating unconscious and implicit cognitive biases to improve healthcare outcomes.

The steps listed above will have a significant effect on bias reduction. When algorithmic bias in healthcare is left unsolved, notable ethical issues linked with data protection and fairness arise. Algorithmic bias within healthcare AI-powered tools and systems can perpetuate and exacerbate ethnic, racial, and gender discrimination. Research has shown that algorithms trained using biased data that do not include critical aspects and statistics from racial minorities, mostly African Americans, is associated with wrong and misleading diagnosis and prognosis. As a result, algorithmic bias within AI-powered programs and tools can lead to disparate and adverse impacts on minorities. Ethnic, racial, and socio-economic minorities continue to receive treatments that do not adequately meet their needs due to misleading algorithms. Socially, algorithmic bias is also a significant cause of stigma, leading to discrimination and stereotyping among patients. Additionally, public stigma compromises proper healthcare providence, leading to coercive treatment and punitive policy-making.

The unwanted and unfair healthcare outcomes minorities face can be addressed by algorithmic bias mitigating policies that work throughout the entire life-cycle of the AI and ML tools. Training data must have an adequate sample size, cover all factual data about a specific group, and be scrutinized for correctness. Similarly, all team members engaged during the algorithm development and validation phases must have medical algorithm checklists that measure fairness, universality, usability, traceability, explainability, and robustness. Close attention to the above attributes will achieve justice and eliminate possible bias in AI and ML tools used in healthcare settings.

References

Agarwal, R., Bjarnadottir, M., Rhue, L., Dugas, M., Crowley, K., Clark, J., & Gao, G. (2023). Addressing algorithmic bias and perpetuating health inequities: An AI bias aware framework. *Health Policy and Technology*, *12*(1), 100702.

Akselrod, O. (2021, July 13). How Artificial Intelligence Can Deepen Racial and Economic Inequities. ACLU. https://www.aclu.org/news/privacy-technology/how-artificial-intelligence-can-deepen-racial-and-economic-inequities

Ahmed, Z., Mohamed, K., Zeeshan, S., & Dong, X. (2020). Artificial intelligence with multi-functional machine learning platform development for better healthcare and precision medicine. Database, 2020(baaa010). https://doi.org/10.1093/database/baaa010

Badal, K., Lee, C. M., & Esserman, L. J. (2023). Guiding principles for the responsible development of artificial intelligence tools for healthcare. *Communications Medicine*, *3*(1), 47.

Beaudouin-Lafon, M., Bødker, S., & Mackay, W. E. (2021). Generative theories of interaction. *ACM Transactions on Computer-Human Interaction (TOCHI)*, *28*(6), 1-54.

Belenguer L. AI bias: exploring discriminatory algorithmic decision-making models and the application of possible machine-centric solutions adapted from the pharmaceutical industry. AI Ethics. 2022;2(4):771-787. doi: 10.1007/s43681-022-00138-8. Epub 2022 Feb 10. PMID: 35194591; PMCID: PMC8830968.

Bernal, J., & Mazo, C. (2022). Transparency of artificial intelligence in healthcare: Insights from professionals in computing and healthcare worldwide. *Applied Sciences*, *12*(20), 10228.

Bernstein, R. S., Bulger, M., Salipante, P., & Weisinger, J. Y. (2020). From diversity to inclusion to equity: A theory of generative interactions. *Journal of Business Ethics*, *167*, 395-410.

Bertl, M., Ross, P., & Draheim, D. (2023). Systematic AI support for decision-making in the healthcare sector: Obstacles and success factors. *Health Policy and Technology*, 100748.

Boal, K. B., & Hooijberg, R. (2000). Strategic leadership research: Moving on. *The Leadership Quarterly*, *11*(4), 515–549.

Čartolovni, A., Tomičić, A., & Mosler, E. L. (2022). Ethical, legal, and social considerations of AI-based medical decision-support tools: A scoping review—*International Journal of Medical Informatics*, *161*, 104738.

Castillo, E. A., & Trinh, M. P. (2019). Catalyzing capacity: Absorptive, adaptive, and generative leadership. *Journal of Organizational Change Management*, *32*(3), 356-376.

Cerdeña, J. P., Plaisime, M. V., & Tsai, J. (2020). From race-based to race-conscious medicine: how anti-racist uprisings call us to act. The Lancet (British Edition), 396(10257), 1125–1128. https://doi.org/10.1016/S0140-6736(20)32076-6

Chowdhury, M. S. (2006). Human behavior in the context of training: An overview of the role of learning theories as applied to training and development. *Journal of Knowledge management practice*, *7* (2), 1–11.

Colquitt, J. A. (2012). Organizational justice. In S. W. J. Kozlowski (Ed.), *The Oxford Handbook of Organizational Psychology*, (Vol. 1, pp. 526–547). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199928309.001.0001

Daneshjou R, Smith MP, Sun MD, Rotemberg V, Zou J. Lack of Transparency and Potential Bias in Artificial Intelligence Data Sets and Algorithms: A Scoping Review. JAMA Dermatol. 2021;157(11):1362–1369. doi:10.1001/jamadermatol.2021.3129

El Bouchefry, K., de Souza, R. S. (2020). Learning in Big Data: Introduction to Machine Learning. In P. Škoda & F. Adam (Eds.), Knowledge Discovery in Big Data from

Astronomy and Earth Observation (pp. 225-249). Elsevier. https://doi.org/10.1016/B978-0-12-819154-5.00023-0

Fejerskov, A. M. (2021). Algorithmic bias and the (false) promise of numbers. *Global Policy*, *12*, 101-103.

Hem MH, Gjerberg E, Husum TL, Pedersen R. Ethical challenges when using coercion in mental healthcare: a systematic literature review. Nurs Ethics. (2018) 25:92–110. 10.1177/0969733016629770

Hoffman, S., & Podgurski, A. (2019). Artificial intelligence and discrimination in health care. Yale J. Health Pol'y L. & Ethics, pp. 19, 1.

Hryciw BN, Fortin Z, Ghossein J, Kyeremanteng K. Doctor-patient interactions in the age of AI: navigating innovation and expertise. Front Med (Lausanne). 2023 Aug 30;10:1241508. Doi: 10.3389/fmed.2023.1241508. PMID: 37711734; PMCID: PMC10498385.

Jackson, M. C. (2021). Artificial intelligence & algorithmic bias: The issues with technology reflecting history & humans. *Journal of Business & Technology Law*, *16*, 299.

Kallus, N., Mao, X., & Zhou, A. (2022). Assessing algorithmic fairness with unobserved protected class using data combination. *Management Science*, *68*(3), 1959-1981.

Khan, B., Fatima, H., Qureshi, A., Kumar, S., Hanan, A., Hussain, J., & Abdullah, S. (2023). Drawbacks of Artificial Intelligence and Their Potential Solutions in the Healthcare Sector. Biomedical materials & devices (New York, N.Y.), 1–8. Advanced online publication. https://doi.org/10.1007/s44174-023-00063-2

Köchling, A., Riazy, S., Wehner, M. C., & Simbeck, K. (2021). Highly accurate, but still discriminatory: A fairness evaluation of algorithmic video analysis in the recruitment context. *Business & Information Systems Engineering*, *63*, 39-54.

Kordzadeh, N., & Ghasemaghaei, M. (2022). Algorithmic bias: Review, synthesis, and future research directions. *European Journal of Information Systems*, *31*(3), 388-409.

Landers, R. N., & Behrend, T. S. (2023). Auditing the AI auditors: A framework for evaluating fairness and bias in high stakes AI predictive models. American Psychologist, 78(1), 36.

Lambert, E. G., Keena, L. D., Leone, M., May, D., & Haynes, S. H. (2020). The effects of distributive and procedural justice on job satisfaction and organizational commitment of correctional staff. *The Social Science Journal*, *57* (4), 405–416.

Lobban, K. (2022). The Anomaly That Is Privacy: Data Privacy Concerns Related to the Rise of Microchip Implants in Humans. https://core.ac.uk/download/541433657.pdf

Luaces, M. A., Galliart, J. M., Mabachi, N. M., Zackula, R., Binion, D., & McGee, J. (2022). Lessons learned from implementing unconscious bias training at an academic medical center. *Kansas Journal of Medicine*, *15*, 336.

Luxton, D. D. (2020). Ethical implications of conversational agents in global public health. *Bulletin of the World Health Organization*, *98* (4), 285.

Lysaght, T., Lim, H. Y., Xafis, V., & Ngiam, K. Y. (2019). AI-assisted decision-making in healthcare: The application of an ethics framework for big data in health and research. *Asian Bioethics Review*, *11*, 299-314.

Martinez-Martin, N., Luo, Z., Kaushal, A., Adeli, E., Haque, A., Kelly, S. S., ... & Milstein, A. (2021). Ethical issues in using ambient intelligence in health-care settings. *The Lancet Digital Health*, *3*(2), e115-e123.

McLeod, S. (2011). Albert Bandura's social learning theory. *Simply Psychology.* London.

Mehrabi, N., Morstatter, F., Saxena, N., Lerman, K., & Galstyan, A. (2021). A survey on bias and fairness in machine learning. *ACM Computing Surveys (CSUR)*, *54*(6), 1-35.

Mello, J., Furlonger, D., Ginsburg, Z., & Kelly, C. (2022). How to reduce cognitive biases in digital business asset initiatives. *ID G00763896*

Mökander, J., Morley, J., Taddeo, M. et al. Ethics-Based Auditing of Automated Decision-Making Systems: Nature, Scope, and Limitations. Sci Eng Ethics 27, 44 (2021). https://doi.org/10.1007/s11948-021-00319-4

Nagoya, R., Bernarto, I., Antonio, F., Pramono, R., Wanasida, A. S., & Purwanto, A. (2021). Exploring intention to enroll university using an extended stimulus-organism-response model. *Academy of Strategic Management Journal*, *20*, 1-12.

Ntoutsi, E, Fafalios, P, Gadiraju, U, et al. Bias in data-driven artificial intelligence systems—An introductory survey. WIREs Data Mining Knowl Discov. 2020; 10:e1356. https://doi.org/10.1002/widm.1356

Panch, T., Mattie, H., & Atun, R. (2019). Artificial intelligence and algorithmic bias: implications for health systems. Journal of Global Health, 9(2).

Pernot-Leplay, Emmanuel and Pernot-Leplay, Emmanuel, EU Influence on Data Privacy Laws: Is the U.S. Approach Converging with the EU Model? (2020). Colorado Technology Law Journal, Vol. 18, No. 1, 2020, Available at SSRN: https://ssrn.com/abstract=3542730

Rumjaun, A., & Fawzia, N. (2020). "Social Learning Theory—Albert Bandura." In Science Education in Theory and Practice: An Introductory Guide to Learning Theory, edited by Ben Akpan and Teresa J. Kennedy, 85–99. Cham: Springer International Publishing, 2020. https://doi.org/10.1007/978-3-030-43620-9_7.

Sanders, D. (2023). The struggle for health: medicine and the politics of underdevelopment. Oxford University Press.

Simon, J., Wong, P. H., & Rieder, G. (2020). Algorithmic bias and the Value Sensitive Design approach. *Internet Policy Review*, *9*(4), 1-16.

Sims Jr., H. P., Faraj, S., & Yun, S. (2009). When should a leader be directive or empowering? How to develop your own situational theory of leadership. *Business Horizons*, *52*(2), 149-158.

Thomasian, N. M., Eickhoff, C., & Adashi, E. Y. (2021). Advancing health equity with artificial intelligence. *Journal of Public Health Policy*, *42*, 602-611.

Thompson, G., & Glasø, L. (2015). Situational leadership theory: A test from three perspectives. *Leadership & Organization Development Journal*, *36*(5), 527-544.

Turner-Lee, N., Resnick, P., & Barton, G. (2019, May 22). Algorithmic bias detection and mitigation: Best practices and policies to reduce consumer harms. Brookings. https://www.brookings.edu/research/algorithmic-bias-detection-and-mitigation-best-practices-and-policies-to-reduce-consumer-harms/?utm_campaign=Brookings+Brief&utm_source=hs_email&utm_medium=email&utm_content=73004492

Vayena, E., Blasimme, A., & Cohen, I. G. (2018). Machine learning in medicine: Addressing ethical challenges. *PLOS Medicine*, *15*(11), e1002689.

Walsh, C. G., Chaudhry, B., Dua, P., Goodman, K. W., Kaplan, B., Kavuluru, R., ... & Subbian, V. (2020). Stigma, biomarkers, and algorithmic bias: Recommendations for precision .behavioral health with artificial intelligence. *JAMIA Open*, *3*(1), 9-15.

West, D.M. & Allen, J., R.(2018). How artificial intelligence is transforming the world. The Brookings Institution. https://www.brookings.edu/articles/how-artificial-intelligence-is-transforming-the-world/

Zarour, M., Alenezi, M., Ansari, M. T. J., Pandey, A. K., Ahmad, M., Agrawal, A., Kumar, R., & Khan, R. A. (2021). Ensuring data integrity of healthcare information in the era of digital health. Healthcare technology letters, 8(3), 66–77. https://doi.org/10.1049/htl2.12008

Zhang, X., Venkatesh, V., & Brown, S. J. (2010). Designing Collaborative Systems for Better Knowledge Management and Team Performance. https://doi.org/10.1109/hicss.2010.137