Marywood University

College of Professional Studies

Healthcare Leaders' Ethical Approaches to AI Bias:
Navigating Algorithmic Bias Across Healthcare Decision-Making

By

Keishalee Shaw

Submitted in Partial Fulfillment of the Requirements for the Degree of
Ph.D. in Strategic Leadership and Administrative Studies

Healthcare Leaders' Ethical Approaches to AI Bias:

Navigating Algorithmic Bias Across Healthcare Decision-Making

Keishalee Shaw

Marywood University

Doctoral Dissertation

March 27, 2025 Date of Approval:	Approved by:	Signatures:	
	Dr. Alan Levine Committee Chair	Signed by: [illegible] F8AE2BC776A14D3...	3/28/2025
	Dr. Matthew Caputo Committee Member	Signed by: [illegible] PhD 1F0C3479C93942B...	3/28/2025
	Dr. Daniel Fallon Committee Member	DocuSigned by: Daniel Fallon A0630637B76D4B2...	3/28/2025
	Mark Friedberg, MD Committee Member	Signed by: Mark Friedberg D7E2F91ABA214A6...	3/28/2025
	Laura Craft Reader	Signed by: Laura Craft EBD049755E67452...	3/28/2025
	Danielle Pelot, JD Reader	Signed by: Danielle Pelot 22C51F845EA6443...	3/28/2025
	Dr. Alexander Dawoody Department Chair	Signed by: Alexander Dawoody C3BA3F008B174AC...	3/28/2025

Healthcare Leaders' Ethical Approaches to AI Bias

Navigating Algorithmic Bias Across Healthcare Decision-Making

The process of addressing algorithmic bias in healthcare decision-making involves traversing intricate ethical and technical terrains where our path alternates between quick and cautious steps while we work both solo and together to encounter challenges that shape us without our awareness. Through ethical insights we discover fair pathways; critical reflection transforms our viewpoints by deepening our understanding; collaborative dialogues create transformative outcomes by uniting diverse perspectives through transparent and accountable processes.

Acknowledgments

Many remarkable people have guided and supported me throughout my academic journey and their invaluable contributions have continuously inspired me.

I would like to thank Dr. Alan Levine for his exceptional mentorship and unwavering support during my dissertation journey. His extensive knowledge of qualitative research and invaluable feedback greatly enhanced my research.

I also want to thank Dr. Mark Friedberg, Dr. Danielle Fallon, and Dr. Matthew Caputo for their invaluable feedback and steadfast support; your contributions have made a lasting impact on my work. I truly appreciate your thorough examination and constructive criticism.

I am grateful to Lee Green and Dr. Yvonne Rowe, whose support and collaboration not only made my research experience productive but also enjoyable. Your friendship was a constant source of motivation.

Enormous thanks go to my family and friends, especially my husband, Steven Shaw; my mother, Elsie Clarke; my sister, Dr. Marshalee Clarke; my mother-in-law Jean Shaw and my children: Katherine, Kristianna, William and Alexander. Their unwavering emotional and practical support played a crucial role in helping me navigate this process. Thank you for your love and belief in me.

I would also like to acknowledge the Dumpling Breadfruit Crew for their ongoing encouragement and support. Your playful yet significant presence helped sustain my focus and energy throughout this challenging yet fulfilling experience.

I sincerely appreciate Shyam Kumar for his continuous support and assistance throughout my research. Your dedication and commitment to my success have been invaluable.

Special thanks also go to Tony Centrella, whose tireless advocacy was instrumental in encouraging my organization to include doctoral programs as part of their professional development benefits.

Finally, I would like to thank every member of my defense committee, including Laura Craft and Danielle Pelot, J.D., for dedicating their time and attention while providing me with valuable critiques throughout the defense process. Your insightful questions and assessments were not only crucial but instrumental in refining my work to its final version.

Table of Contents

Abstract

Artificial intelligence (AI) and machine learning (ML) are transforming healthcare practices. They aid in diagnosis and efficiently handle administrative tasks. However, algorithmic bias has significantly affected fairness and equity in healthcare delivery. This research examines the methods used by healthcare leaders to handle algorithmic bias within AI-based decision-making frameworks. The study uses interpretative phenomenological analysis (IPA) to investigate healthcare leaders' views on the ethical, organizational, and technical barriers they face while managing bias during data collection and analysis and using this information for healthcare decision-making. Algorithmic bias due to imbalanced training, data transparency, and a defective legislative model increases healthcare disparities between minority groups, including misdiagnoses, unequal distribution of resources, and loss of confidence in medical technology.

This research advances broader AI governance discussions by outlining areas where leaders can identify and address the technical and ethical complexities of AI implementation. The study also bridges theory and empirical insights to provide suggestions on improving fairness, transparency, and accountability. To mitigate bias, leaders recommend four key measures: diverse datasets, continuous observation, multidisciplinary collaborations, and cultural competency training protocols. Leaders believed that data should be ethically protected under strong regulatory frameworks, supervised by those with legitimate oversight. They suggest continuous monitoring, diversity in algorithm development teams, and patient-centered design approaches. The study demonstrates the need for a complete solution to algorithmic bias that includes practical computational methods, ethical institutional guidelines, and cultural development strategies. Organizations that implement transparent practices and promote inclusivity while holding themselves accountable will create effective strategies to minimize bias and achieve fair outcomes.

Keywords: Algorithmic bias, data control, healthcare leadership, ethical governance, interpretative phenomenological analysis, healthcare equity, AI transparency, patient-centered design, machine learning.

List of Tables

List of Figures

Chapter 1
The Problem and Its Setting

Introduction

Within the healthcare sector, algorithms powered by artificial intelligence (AI) and machine learning (ML) are increasingly crucial in making decisions related to patient care and resource allocation, with AI algorithms focusing on broader decision-making capabilities and ML algorithms emphasizing pattern recognition and predictive analytics. Both AI and ML are essential for enhancing healthcare quality and facilitating its transition from traditional, manual processes to more automated and efficient systems (Ahmed et al., 2020).

AI entered the healthcare field in the mid-1950s with early efforts to create clinical expert systems—computerized decision-making tools designed to emulate human reasoning and assist or replace doctors in diagnosing and treating patients. This era saw the development of pioneering systems like MYCIN and INTERNIST in the 1970s, marking significant progress in medical AI research (see Appendix A for a detailed historical breakdown). Throughout the late 1970s and early 1980s, several other expert systems were developed, each focusing on specific medical diagnoses and utilizing advanced computational models. However, challenges such as system performance limitations and reliance on expensive hardware hindered widespread adoption until advancements in the mid-1990s paved the way for more practical applications, including electronic medical records and telemedicine initiatives, laying the foundation for future AI-driven healthcare innovations (Bhargava & Sharma, 2021). AI-driven healthcare systems enhance healthcare delivery through improved diagnosis, treatment planning, patient monitoring, and administrative efficiency. These systems leverage algorithms and data analysis to improve efficiency, accuracy, and outcomes in healthcare (Dlugatch et al., 2024).

Machine learning (ML), a subset of AI, utilizes algorithms and data to mimic human learning and improve accuracy (IBM, n.d.). In 1959, Arthur Samuel, a pioneering computer scientist in artificial intelligence, coined "machine learning" as the term for computers' ability to learn autonomously, without explicit programming. Alan Turing's landmark paper, *Computing Machinery and Intelligence*, in 1950 established a yardstick for assessing machine intelligence, stipulating that machines must demonstrate responsiveness indistinguishable from humans.

Applying this concept to healthcare, machine learning involves computers assimilating past medical data to anticipate future outcomes, akin to a physician learning from patient cases to improve diagnosis accuracy. Tom M. Mitchell's technical definition from 1997 further delineates machine learning as a computer program's ability to enhance task performance (T) through experience (E), as measured by a performance metric (P). For instance, in healthcare, a machine learning algorithm analyzing patient records (E) to predict disease progression (T) and achieving increasingly accurate predictions (P) over time epitomizes this process (Mitchell, 1977).

As Hoffman & Podgurski (2020) have pointed out, however, algorithmic discrimination and bias is not uncommon in healthcare, manifesting in race-based adjustments or failing to identify candidates for high-risk care management. Healthcare bias can encompass various complicated and interrelated issues. Algorithmic bias, for example, refers to a consistent mistake in how computers predict outcomes. This mistake can happen because of errors in the code, choosing the wrong model, using the wrong criteria for improvement, or ignoring certain data (Djap, 2022). In this study, algorithmic bias refers to the systematic and unfair biases that can be present in algorithms, leading to discriminatory outcomes or decisions. These biases can stem from human biases embedded in the data used to train the algorithms, from the design choices made in developing the algorithms, or from the specific context in which they are applied.

This study presupposes the necessity of multifaceted approaches to address the ethical and practical challenges posed by AI in healthcare. By drawing on a diverse range of literature, including discussions on accountability frameworks, transparency, fairness, and interdisciplinary collaboration, the study implicitly acknowledges the complexity of the issues at hand and the need for comprehensive, interdisciplinary solutions.

The rationale for selecting interpretative phenomenological analysis (IPA) as the methodological framework for this research lies in its suitability for exploring the nuanced and subjective experiences of healthcare leaders grappling with algorithmic bias in healthcare decision-making. IPA emphasizes the exploration of individuals' lived experiences and perceptions, making it particularly apt for capturing the rich and contextualized insights sought in this study. By employing IPA, this research aims to uncover not only the surface-level challenges but also the deeper meanings and interpretations healthcare leaders attach to their encounters with algorithmic bias. Additionally, IPA allows for a flexible and iterative approach, aligning well with the iterative nature of sense-making and decision-making processes within healthcare leadership.

The theoretical and conceptual frameworks drawn from the works of Larson et al. (2021), Helm et al. (2020), Barton et al. (2023), and Ahmed et al. (2020) provide a solid foundation for understanding the broader regulatory, ethical, and transparency considerations inherent in the study's focus. By integrating these frameworks with the IPA methodology, I aim to not only explore the lived experiences of healthcare leaders but also to contextualize these experiences within the broader landscape of fairness, transparency, and ethical conduct in AI-driven healthcare systems. Thus, the choice of IPA is strategically positioned to facilitate a comprehensive exploration of the research questions, ultimately aiming to contribute valuable insights and recommendations for promoting fairness, transparency, and accountability in healthcare decision-making processes influenced by AI algorithms.

The purpose of this study is to explore and understand the experiences of healthcare leaders in addressing algorithmic bias within the realm of data collection, analysis, and utilization in healthcare decision-making. The use of a qualitative, interpretative phenomenological approach allows for several key advantages in the context of researching healthcare leaders' experiences with algorithmic bias in healthcare decision-making. The IPA methodology provides insights into a particular situation and perspective, as opposed to a generalizable, broad overview of an issue, via investigating lived experiences of healthcare leaders. Healthcare leaders encompass a spectrum of roles, from executives and administrators to managers and directors, spanning across hospitals, clinics, pharmaceutical firms, insurance entities, governmental bodies, and beyond. A healthcare leader within this study refers to someone who occupies a position of authority and directs the management and strategic operations of healthcare organizations or teams. Leadership within clinical settings encompasses administrative leaders and clinicians including attending physicians who serve as leaders in healthcare entities or teams. These entities may be connected through joint ownership, management, or affiliation via shared ownership, contractual agreements, or unified management. The primary goal of such a system is to offer comprehensive care encompassing both primary and specialty services (AHRQ, 2023). For this purpose of this study the term "healthcare system" will encompass all the organizations, institutions, resources, policies, and practices involved in providing healthcare services to individuals and populations. It includes healthcare providers, hospitals, clinics, insurance providers, government agencies, and other stakeholders involved in promoting and maintaining public health and well-being.

Healthcare leaders play a crucial role in healthcare information technology (HIT) implementation, yet many leaders are unaware of their responsibilities in this area; with the evolving healthcare landscape and technological advancements, the role of healthcare leaders has expanded to include proficiency not only in clinical health services and management but also in health information technologies, necessitating further research into the impact of leaders at all levels on HIT implementation (Laukka, E., et al., 2020). As Dr. Robert Baginski, the program director of Northeastern University's Doctor of Medical Science (DMSc) in Healthcare Leadership program, articulates, "Healthcare should be looking to the future, toward solutions where we can provide the best and the most healthcare to those who need it most—regardless of insurance status, access to money, or where they live" (Joubert, 2023).

The study operates under the assumption that healthcare leaders wield significant influence in shaping the trajectory of AI-driven healthcare practices. It assumes that these leaders are pivotal actors in the formulation of policies and strategies aimed at mitigating algorithmic bias and fostering ethical considerations in healthcare decision-making processes. This assumption underscores the critical role of leadership in navigating the complex intersection of technology and ethics within healthcare systems.

Finally, according to the *International Encyclopedia of Human Geography*, second edition (2020), lived experience is a "term used throughout phenomenology to denote the immediate and subjective everyday lived experience of an individual or group, through which meaning is made." In this study, lived experiences refer to the subjective and personal experiences of healthcare leaders, based on their unique circumstances, backgrounds, and interactions with the world around them. These experiences shape individuals' perspectives, beliefs, and understanding of various aspects of life. This definition will serve to focus the current research parameters around the experiences and opinions of the participants. Having introduced the broader landscape of AI and ML within healthcare, the following section elaborates on the theoretical foundations that underpin this research, focusing particularly on machine learning theory.

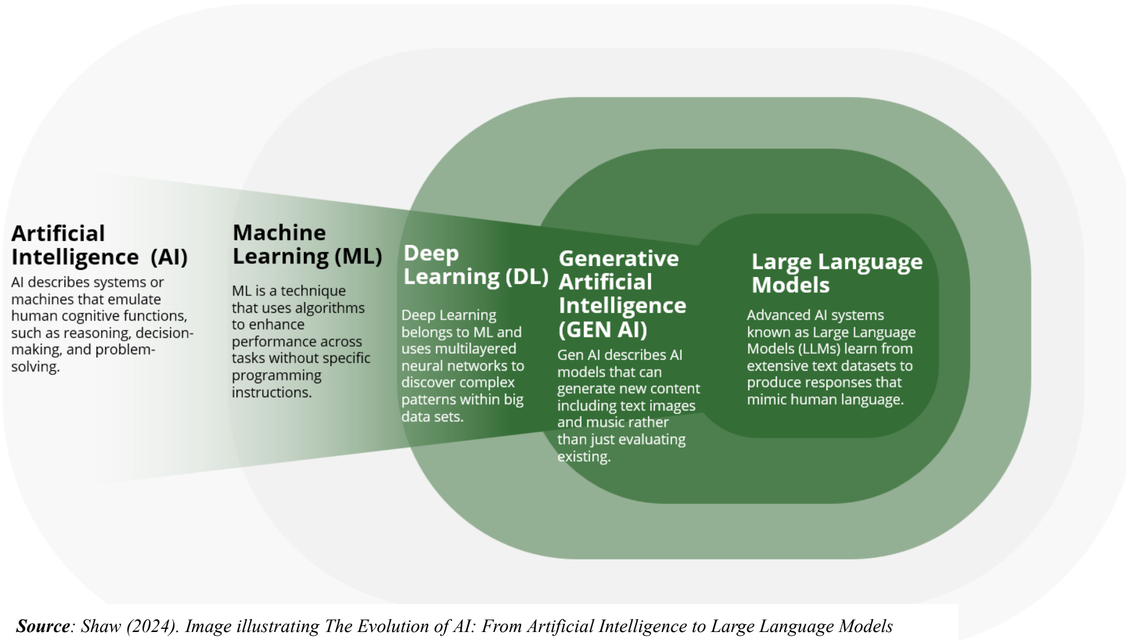
Theoretical Framework

The theoretical framework of this study is structured to offer a comprehensive understanding of the intricate issues surrounding algorithmic bias in healthcare, drawing upon the foundations of machine learning and artificial intelligence. However, at its essence, the study investigates the fundamental principles of machine learning theory (Singh & Sinha, 2022). These principles enable computers to autonomously recognize patterns and make decisions from data, devoid of explicit programming instructions.

Artificial Intelligence (AI) encompasses machines replicating human cognitive abilities, utilizing patterns to match expert systems (*Difference between AI vs Machine Learning vs Deep Learning*, n.d.). AI is widely employed across various industries, including retail, finance, education, and healthcare. Three main types of AI support business needs: data collection and analysis, business automation, and engagement with customers or employees.

Figure 1

Artificial intelligence ⇨ Machine Learning ⇨ Deep Learning ⇨ Generative Artificial *I*ntelligence ⇨ Large Language Models



Source*: Shaw (2024). Image illustrating The Evolution of AI: From Artificial Intelligence to Large Language Models*

Figure 1 above depicts how the concepts of AI, ML, and deep learning are interconnected to deal with vast amounts of data. Using modern technology, organizations can constantly collect, analyze, and utilize information, choosing between AI, ML, and deep learning for every specific case, with each offering unique advantages. For example, ML enables machines to learn from patterns without human intervention. It is extensively utilized in sectors like banking, with companies like Apple, Amazon, and Microsoft employing ML in their services and products, such as virtual assistants and recommendation systems (Bini, 2018). Deep learning, on the other hand, is a further subset of ML. It involves machines learning without relying on pre-defined patterns, requiring large sets of labeled data and extensive training. It leverages neural networks, mimicking the human brain's decision-making process. Deep learning has applications ranging from pattern recognition to medical diagnoses, exemplified by Google's AlphaGo defeating a human champion in the complex game of Go (*Difference between AI vs Machine Learning vs Deep Learning,* n.d.). The overarching goal is to develop cutting-edge software solutions to address business challenges effectively.

As explained above, machine learning is often seen as part of AI. Both involve using experiences to learn and find patterns in data, similar to how humans and animals learn. However, they differ in their goals. Traditional AI tries to copy intelligent behavior, while machine learning uses computers to enhance human intelligence, often handling tasks humans cannot (Nabi, 2018). For example, in healthcare, machine learning helps by analyzing large databases, spotting patterns a human might miss.

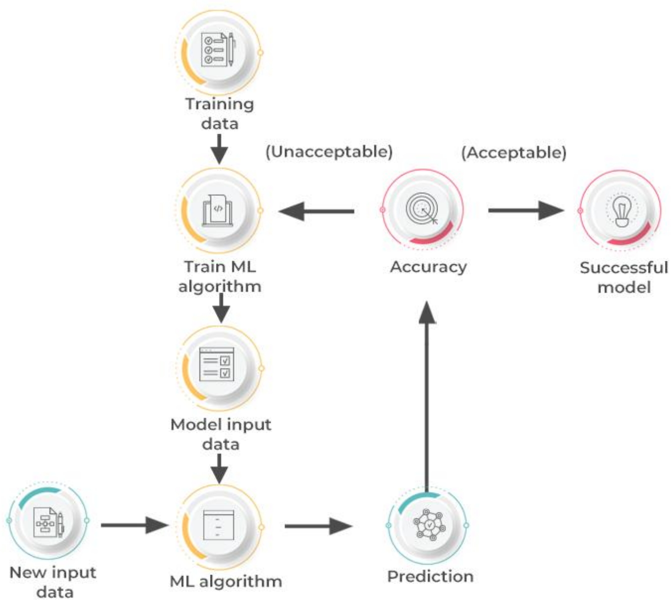
Machine learning algorithms are molded on a training dataset to create a model (see Appendix A). Figure 2 shows a high-level use case scenario. A grasp of machine learning theory is essential for grasping the nuances of algorithmic bias—its origins, manifestations, and potential remedies. By anchoring this research within machine learning theory, we seek to establish a robust framework for examining the experiences of healthcare leaders as they navigate the challenges presented by algorithmic bias.

Figure *2*
How Machine Learning Transforms Data into Decisions

Source: Rojewska, K. (2023). How Machine Learning Works. Qtravel.aiBlog. Qtravel.ai Blog

Figure 2 above provides a step-by-step view of how machine learning processes raw data into meaningful decisions by following stages of collection, preparation, analysis, pattern recognition, prediction, and action. The initial phase of machine learning involves gathering various data types (photos, text, numbers) which must be cleaned to correct mistakes and maintain accuracy—an optional yet advisable step to enhance outcomes. Through data analysis the machine learning model identifies the underlying structure and element relationships while seeking hidden patterns and links. These identified patterns enable the model to forecast outcomes from previously unseen data. The system uses predictions to make decisions and perform actions such as identifying spam emails and sorting them into the correct folder.

Figure ***3***
How Machine Learning Works



Source: **Figure 3** How machine learning works [Illustration]. From "How Machine Learning Works," by V. Kanade, 2022, *Spiceworks*.

Figure 3 comprises several interconnected components. When new input data is introduced to a trained machine learning (ML) algorithm, it utilizes the developed model to make predictions. Imagine you are training an AI system within a healthcare insurance company to recognize and process claims. Using figure 3 above, let us walk through how it works using a healthcare example.

Step One: Training Data

- You provide the system with a diverse set of examples of healthcare claims, including both approved and denied cases.

Step Two: Training the ML Algorithm

- The system practices evaluating these claims, learning to distinguish between valid and invalid ones. If the system does not identify enough claims correctly during training, you refine its learning by adjusting the algorithm.

Step Three: Achieving Accuracy

- As the system becomes proficient at evaluating claims, achieving high accuracy, it evolves into a "successful model." This model can now predict whether new claims are valid or not.

Step Four: New Input Data and Predictions

- When you present the system with new, previously unseen claims (referred to as "new input data"), it applies its learned knowledge to independently assess their validity. The system's informed evaluation of these claims is called a "prediction."

This process enables the AI system to autonomously determine the validity of healthcare claims based on its learned patterns, which in turn allows for new input data to be introduced to the trained ML algorithm. This will help to develop the model to make predictions going forward (Kanade, 2022). Given these ethical considerations, understanding fairness becomes essential; thus, the subsequent section explores fairness as a fundamental ethical component in healthcare AI.

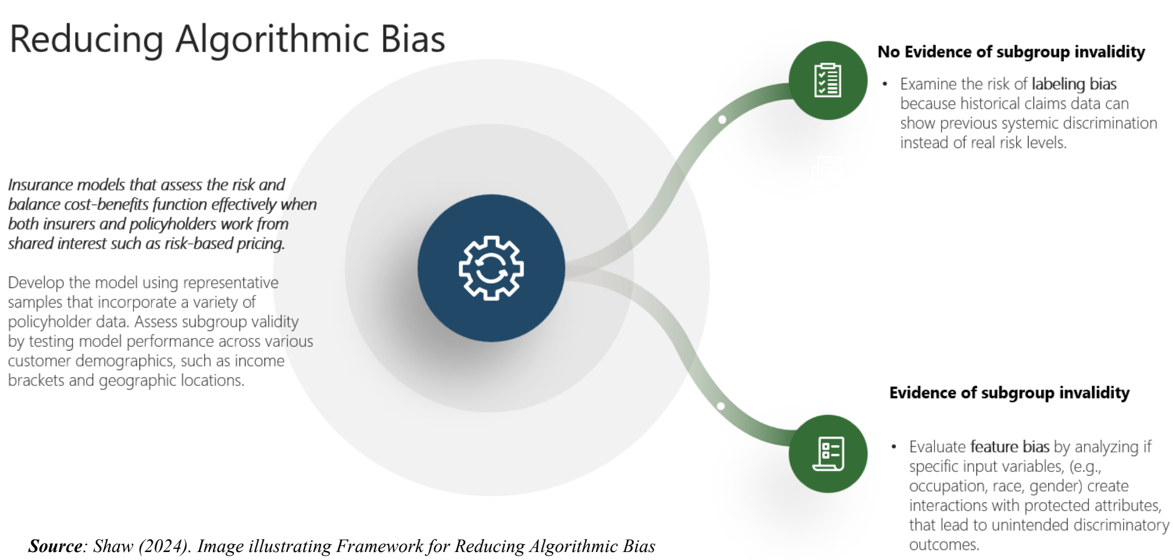
Conceptual Framework

The conceptual framework of this study targets bias reduction in AI-based healthcare systems. Addressing bias in healthcare AI requires an extensive research approach which must incorporate technical aspects along with ethical principles and regulatory standards as well as societal impacts. The core of our framework is Augmented Intelligence in Machine Learning (AIML), which proposes a new paradigm that supports human–machine cooperation instead of

AI systems supplanting human decision-making. The partnership between human expertise and machine intelligence brings together multiple healthcare domains to achieve transparent and accountable decision-making processes which result in better patient care. The AIML framework develops from the research conducted by Larson and his team (2021), Helm et al. (2020), and Barton et al. (2023). The AIML framework builds on existing studies to underline the necessity for legal standards in conjunction with algorithmic fairness. This fairness concept requires healthcare resources and opportunities to be distributed equally among populations without regard for racial, gender, or socioeconomic background differences.

In the case of healthcare and based on these principles, Figure 4 and Figure 5's conceptual framework concentrates on two main aspects to diminish algorithmic bias within insurance decision-making models is presented. Figure 4 focuses on Reducing Algorithmic Bias and Figure 5 focuses on Reducing Algorithmic Unfairness.

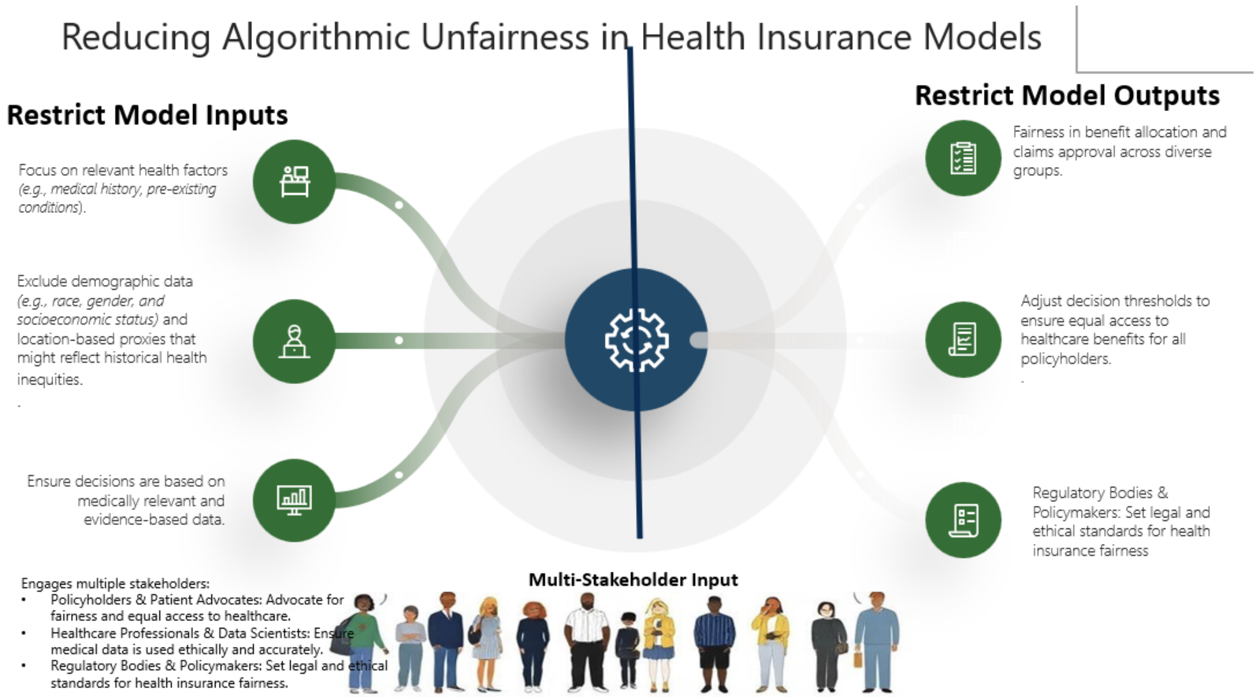
Figure *4*
Reducing Algorithmic Bias



***Source**: Shaw (2024). Image illustrating Framework for Reducing Algorithmic Bias*

Source*: Shaw (2024). Image illustrating Framework for Reducing Algorithmic Unfairness in Health Insurance Models*

Figure 5
Reducing Algorithmic Unfairness



Source: *Shaw (2024). Image illustrating Framework for Reducing Algorithmic Unfairness in Health Insurance Models*

For risk assessment models where insurers and policyholders share aligned interests, bias reduction depends on using representative samples during development and evaluating subgroup validity across demographic segments. When no validity issues emerge, attention shifts to potential labeling bias where historical claims data may reflect past discrimination rather than actual risk; when subgroup invalidity appears, feature bias analysis becomes essential to identifying how variables like occupation might interact with protected attributes to create discriminatory outcomes.

For models controlling resource distribution and eligibility, we recommend two complementary approaches: restricting inputs to proven causal risk factors while eliminating both direct demographic identifiers and indirect proxies, and evaluating and constraining outputs by analyzing service distribution across groups and potentially adjusting decision thresholds or implementing fairness constraints. Successful implementation requires multi-stakeholder collaboration between policyholders, consumer advocates, actuaries, data scientists, and regulatory bodies to ensure insurance models effectively integrate ethical and legal standards throughout the risk assessment process.

Purpose Statement

The research for this paper builds upon existing literature stressing the importance of legal standards, transparency, and fairness. The purpose of this qualitative study is to explore and understand the experiences of healthcare leaders in addressing algorithmic bias within the realm of data collection, analysis, and utilization in healthcare decision-making. The research draws upon insights from regulatory oversight, transparency, fairness, and accountability frameworks proposed by Hall et al. (2015), Aisha Shalash et al. (2022), and Ahmed et al. (2020), with the aim of contributing to the promotion of fairness, transparency, and ethical conduct in AI-driven healthcare systems, using an interpretative phenomenological approach.

Central Research Question

What are the lived experiences of healthcare leaders in addressing algorithmic bias in healthcare decision-making, and how do these experiences inform strategies for promoting fairness and accountability in AI-driven healthcare systems?

Research Sub-Questions

1. Perception of Algorithmic Bias: How do healthcare leaders perceive algorithmic bias within the context of healthcare decision-making, and what factors influence these perceptions?
2. Strategies for Bias Mitigation: What strategies do healthcare leaders employ to mitigate algorithmic bias in healthcare decision-making processes, and what challenges do they encounter in implementing these strategies?

3. Challenges in Ethical Considerations: What are the key challenges healthcare leaders face in navigating ethical considerations, patient-centered design, and collaborative decision-making in addressing algorithmic bias in healthcare systems?

By addressing these sub-questions, this study aims to provide insights into algorithmic bias's complexities in healthcare decision-making and offer recommendations for fostering fairness, transparency, and accountability in AI-driven healthcare systems.

Key Terminologies Defined

In this study, key terms are clarified as follows based on their application.

Accountability involves being responsible for one's actions, decisions, and their consequences (Krautscheid, 2014). For this study, accountability entails ensuring that healthcare providers, institutions, and systems are held responsible for delivering high-quality care, adhering to ethical standards, and addressing any errors or shortcomings.

AI-driven Healthcare Systems utilize artificial intelligence to simulate human intelligence in tasks such as analyzing, presenting, and understanding complex medical and healthcare data to enhance various aspects of healthcare delivery, including diagnosis, treatment planning, patient monitoring, and administrative tasks. These systems leverage algorithms and data analysis to improve efficiency, accuracy, and outcomes in healthcare, especially by offering new methods to diagnose, treat, or prevent diseases (Dlugatch et al., 2024).

Algorithmic Bias refers to a consistent mistake in how computers predict outcomes. This mistake can happen because of errors in the code, choosing the wrong model, using the wrong criteria for improvement, or ignoring certain data (HealthITSecurity, 2023). In this study algorithmic bias refers to the systematic and unfair biases that can be present in algorithms, leading to discriminatory outcomes or decisions. These biases can emerge from the data used to train the algorithms, the design of the algorithms themselves, or the context in which they are applied.

Bias Mitigation is the process of identifying, addressing, and eliminating biases that may be present, particularly in situations where racial bias may impact patient care (Barber et al., 2023). In this study, bias mitigation involves the process of identifying, acknowledging, and reducing biases present in algorithms or decision-making processes. This can include techniques such as data preprocessing, algorithmic adjustments, diversity in data collection, and ongoing monitoring to ensure fair and equitable outcomes.

Decision-making refers to the process of selecting a course of action or choice among several alternatives based on evaluation, analysis, and judgment. In healthcare, decision-making encompasses a wide range of activities, including clinical decision-making by healthcare providers, administrative decisions by healthcare organizations, and policy decisions by policymakers (Dlugatch et al., 2024)

Fairness in "healthcare is a multidimensional concept that includes the equitable distribution of resources, opportunities, and outcomes among diverse patient populations" (Ueda, 2024). Fairness is often referred to as the quality of being just, equitable, and impartial. In the context of healthcare and for this study, fairness entails ensuring equal access to healthcare services, resources, and opportunities for all individuals, regardless of factors such as race, gender, socioeconomic status, or other characteristics.

Healthcare Leaders play a crucial role in healthcare information technology (HIT) implementation, yet many leaders are unaware of their responsibilities in this area; with the evolving healthcare landscape and technological advancements, the role of healthcare leaders has expanded to encompass proficiency not only in clinical health services and management but also in health information technologies, necessitating further research into the impact of leaders at all levels on HIT implementation (Laukka et al., 2020). In this study healthcare leaders refer to individuals who hold positions of authority or influence within their healthcare organization/industry. They may include executives, administrators, policymakers, and professionals who play a significant role in shaping the direction, policies, and practices of healthcare organizations or systems.

Healthcare System. The Agency for Healthcare Research and Quality gathered diverse definitions from various Centers of Excellence to clarify the concept of a health system. Essentially, a health system refers to a structured network that includes multiple healthcare entities, such as hospitals, physician practices, and other facilities. These entities may be connected through joint ownership, management, or affiliation via shared ownership, contractual agreements, or unified management. The primary goal of such a system is to offer comprehensive care encompassing both primary and specialty services (AHRQ, 2023). For the purpose of this study the term "healthcare system" encompasses all the organizations, institutions, resources, policies, and practices involved in providing healthcare services to individuals and populations. It includes healthcare providers, hospitals, clinics, insurance providers, government agencies, and other stakeholders involved in promoting and maintaining public health and well-being.

Lived Experiences According to *International Encyclopedia of Human Geography* (second edition) (2020). Lived experience is a "term used throughout phenomenology to denote the immediate and subjective everyday lived experience of an individual or group, through which meaning is made." In this study lived experiences refer to the subjective and personal experiences of individuals based on their unique circumstances, backgrounds, and interactions with the world around them. These experiences shape individuals' perspectives, beliefs, and understanding of various aspects of life.

Patient-centered Design The significance of Human-Centered Design (HCD) in prioritizing user needs and satisfaction can be described as similar to the Patient-centered design. HCD involves understanding users, tasks, and environments; involving users throughout design and development; driving design with user-centered evaluation; iterating the process; addressing the entire user experience; and incorporating multidisciplinary skills and perspectives (Meskó & Topol, 2023). For this study Patient-centered design is an approach to healthcare design and delivery that prioritizes the needs, preferences, and experiences of patients. It involves actively involving patients in the design process, tailoring healthcare services to meet individual patient needs, and emphasizing empathy, communication, and collaboration between patients and healthcare providers.

Delimitations

Certain delimitations constrain the scope of this study. First, in the case of geographic boundaries, the focus is on healthcare leaders' perspectives along the East Coast, primarily within Maryland and Massachusetts. Second, the targeted demographic group of healthcare leaders and executives allows the researcher to narrow the field of inquiry to those most directly involved in decision-making processes related to data within healthcare settings. Third, the study's population constraints specifically capture leaders within healthcare entities interfacing with data regularly, emphasizing a qualitative approach to gather nuanced perspectives. Additionally, the study's methodology confines data collection to qualitative methods and contemporary literature, aligning with the aim of comprehensively exploring leadership dynamics in the context of Generative AI within the healthcare domain. Fifth, the scope of analysis for this paper is about the role leadership plays when it comes to Generative AI/ML and its use to provide service to consumers, as opposed to non-AI solutions such as traditional human-based services or rule-based algorithms (Zhang et al., 2022). Lastly, this study recognizes the temporal limitations, with a research period of three months for interviews and a focus on literature published within the past five to six years, ensuring relevance and timeliness of the analysis.

Assumptions

The study assumes that participants will offer candid and insightful reflections on their encounters with algorithmic bias. By assuming participants' willingness to share their experiences and challenges openly, the study anticipates rich qualitative data that can inform effective strategies for addressing bias in AI-driven healthcare decision-making. This assumption underscores the importance of fostering an environment conducive to open dialogue and reflection among stakeholders.

Significance of Study

The significance of the study lies in its potential to influence/reshape the landscape of healthcare decision-making by addressing the pervasive issue of algorithmic bias and promoting ethical considerations in AI-driven systems. Healthcare leaders, policymakers, researchers, and practitioners are key stakeholders who stand to benefit from the insights gleaned from this research. By focusing on the lived experiences and perspectives of healthcare leaders, the study aims to provide valuable insights that can inform the development of targeted interventions, guidelines, and best practices aimed at fostering fairness, transparency, and accountability in AI-enabled healthcare systems.

One crucial aspect highlighted in the study is the need for regulatory oversight to govern the development, deployment, and use of health technologies. Larson et al. (2021) stress the importance of legal standards to ensure patient privacy, ethical conduct, and algorithm validation, emphasizing the role of regulations in upholding ethical principles in healthcare AI. The study further underscores the importance of transparency and fairness in the decision-making processes, as highlighted by News (2023), advocating for transparent AI use to minimize bias.

Ethical issues, including bias mitigation, patient-centered design, and privacy concerns, are also central to the study's significance. Chen et al. (2023) discusses techniques to reduce biases and ensure equality within healthcare services, while Rines et al. (2022) emphasizes the importance of integrating patient perceptions into the design phase of AI-powered health systems to improve outcomes and promote personalized, inclusive, and equitable healthcare delivery. Furthermore, Wang et al. (2023) examine the ethical implications of using conversational AI models in clinical environments, focusing on issues such as patient privacy, consent, and potential biases in outcomes.

Overall, the study advocates for a collaborative and interdisciplinary approach to addressing algorithmic bias and techno-ethical issues in healthcare AI. Aquino et al. (2023) stress the importance of multidisciplinary collaboration involving clinicians, data scientists, ethicists, and policymakers to tackle these challenges effectively. Additionally, the study highlights the significance of qualitative research methods in navigating the complexities of AI technology adoption in healthcare institutions, as discussed by Pathak et al. (2013), underscoring the importance of understanding the perspectives and cultural influences shaping AI-driven healthcare systems.

This research also sheds light on the importance of cost management and implementation strategies in ensuring the widespread adoption and sustainability of AI-driven healthcare solutions. Avanceña & Prosser (2021) discuss cost-effectiveness analysis as a tool for evaluating healthcare interventions and optimizing resource allocation for AI implementation, while Celi et al. (2022) explore global healthcare disparities and sources of bias created by AI, advocating for preemptive strategies to counter bias and ensure equity in healthcare delivery. Overall, the study contributes to the broader discourse on augmenting human intelligence with machine learning technologies to optimize healthcare outcomes while upholding ethical principles and patient-centered care.

Chapter 2
The Literature Review

Introduction

In recent years, the intersection of machine learning (ML) and healthcare has seen a surge of interest in innovation. The advent of augmented intelligence, a term coined to emphasize the collaborative potential of human expertise and machine capabilities, has promised transformative advancements in healthcare decision-making processes. Augmented intelligence in machine learning (AIML) refers to the integration of human expertise and machine intelligence to enhance decision-making processes. However, with any technological integration in sensitive domains such as healthcare, considerations surrounding costs, accountability, and ethics loom large.

The following literature review on the impact of technology on healthcare begins with a detailed exploration of the research strategy employed. It then explores key technological advancements, particularly focusing on machine learning and its applications within healthcare. Subsequently, the review discusses the persistent issue of algorithmic bias, covering its types, impacts, and ethical considerations, including fairness, accountability, and privacy concerns. Moreover, it addresses multidisciplinary approaches, referring to methods or strategies that combine knowledge and techniques from different fields including medicine, ethics, technology, and policy alongside the integration of technological advancements and ethical principles. Additionally, the review examines the role of people-powered AI, cost management strategies, and the significance of patient-centered design in implementing technology effectively within healthcare systems. Finally, the definition and basis of potential mitigation methods from various fields are given.

Literature Search Strategy

The strategy adopted for this review involved a systematic and comprehensive approach aimed at identifying pertinent studies, articles, and reports within the realm of augmented intelligence in healthcare decision-making. Initially, a combination of electronic databases including PubMed, DOI, IEEE Xplore, Scopus, and Google Scholar were queried to ensure broad coverage across medical, engineering, and interdisciplinary fields. Boolean operators such as "AND," "OR," and "NOT" were employed to refine the search queries and enhance search precision.

A manual search of key journals and conference proceedings in the fields of healthcare informatics, machine learning, medical ethics, and health policy was conducted to supplement electronic database searches. Additionally, reference lists of identified articles and reviews were scrutinized to uncover potentially overlooked sources to ensure a comprehensive examination of the literature.

An Overview of Machine Learning

Sadiku et al., (2021) explains how augmented intelligence involves the application of AI technologies to improve human cognitive abilities and decision-making talents. Unlike typical AI, augmented intelligence aims to enhance human intelligence. AI assimilation incorporates smart algorithms and tools within human workflows to multiply cognitive capabilities and stimulate performance levels in different sectors, for instance, medicine.

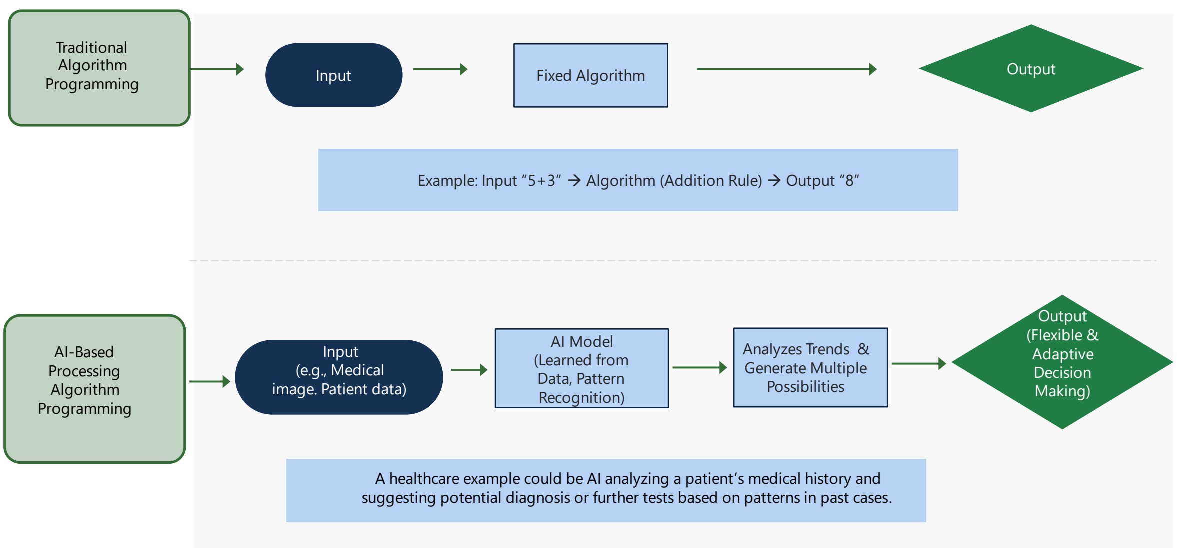
Helm et al. (2020) illustrate how machine learning algorithms autonomously identify patterns from data, integrating clinical expertise with technological capability to enhance predictive accuracy in healthcare settings. Machine learning algorithms are capable of recognizing patterns and connections between data points, leading to computers becoming more efficient through experience (Helm et al., 2020). In the healthcare industry, machine learning algorithms are adapted for different purposes, for instance, the diagnosis of uncommon diseases, personalized treatment goals, and patient image analysis.

ML is also used in decision-making. According to Shinkunas et al. (2020), healthcare decision-making is the process of appraising information regarding medical practice, the attitude of the patient, and the programs of treatment to arrive at the best decision for the patient (Shinkunas et al., 2020). It implies an evaluation of evidence-based procedures, clinical guidelines, and ethical concerns to enhance patient outcomes to a high ratio while delivering healthcare of the best quality. Healthcare decision-making concerns a variety of stakeholders, such as medical professionals, patients, caregivers, and politicians, who work together to determine the most appropriate treatment plan for every individual patient. Because healthcare decision-making directly impacts patient well-being, it is critical to identify and mitigate algorithmic bias.

Algorithmic Bias

Algorithms function like recipes that tell a computer how to solve a problem step by step. They function as a collection of sequential instructions which guide the process of solving a problem or executing a task. The algorithm functions by taking input data, executing multiple operations on it, and then generating the output. Think of it like a recipe: The desired result becomes possible when you execute steps sequentially. Algorithms are used in math, computer science, and data organization. For example, insurance companies use algorithms to determine the cost of medical procedures based on factors like a patient's age, medical history, and the type of treatment needed (Collins, 2024). Figure 6 below depicts a simple breakdown of *Traditional Algorithm Programming vs AI-Based Processing*: Traditional computer programming works by processing inputs through a defined series of steps called algorithms to deliver consistent results. The program consistently outputs "8" when processing the input "5 + 3" because it operates according to predetermined rules.

Figure 6
Traditional Algorithm Programming vs AI-Based Processing



Source*: Shaw (2025). Traditional Algorithm Programming vs AI Based Processing*

On the other hand, however, with AI, the process is different. AI systems require input data such as medical images or patient information, but they do not operate solely based on fixed programming rules. The system depends on a data-trained model that detects patterns to inform its decision-making process. When an AI system evaluates a patient's medical history to determine possible health conditions it examines both specific details and patterns discovered from numerous comparable situations. Based on its knowledge gained through learning, AI generates multiple potential diagnoses and can suggest additional tests instead of delivering a single static solution. AI systems now have the ability to adjust to complex scenarios because they provide adaptable solutions which surpass traditional systems that rely on fixed rules.

Hoffman & Podgurski (2020) have identified algorithmic discrimination as a recurring issue in healthcare. Algorithmic discrimination refers to preferences that emerge from the design and implementation of AI algorithms. Collectively, these shortcomings produce 'algorithmic bias.' "Algorithmic bias in health systems can exacerbate existing inequities in socioeconomic status, race, ethnic background, religion, gender, disability, or sexual orientation, ultimately amplifying disparities in health outcomes" (Panch, Mattie, & Atun, 2019). According to Yves 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. Biases in healthcare are generally tied to data, outcome, patient attributes, clinical decision-making, or user bias.

AI and ML systems in healthcare rely heavily on data for training and decision-making, primarily through a two-step process. First, in data collection, relevant data is gathered from various sources for analysis and decision-making purposes. Then, data analysis, defined as the systematic examination and interpretation of data, is used to uncover meaningful insights and patterns. The application of data-driven insights and findings to inform decision-making processes and improve patient care in healthcare settings collectively is known as data utilization.

Types of Bias

In healthcare, bias falls roughly into five categories: data bias, outcome bias, patient bias, clinical bias, and user bias. Each plays a role in healthcare discrimination and thus warrants a definition and explanation as part of this study. First, 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 adjust and improve when used on individuals from other demographics.

Outcome bias occurs when AI and ML systems are evaluated based on specific outcomes or performance metrics that may not capture the systems’ full impact on healthcare disparities. “AI Bias is when the output of a machine-learning model 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. An example of this is UnitedHealthcare’s use of AI to handle claim denials which drew criticism due to its high error rate causing possible discriminatory results. UnitedHealthcare’s AI’s pursuit of cost savings resulted in treatment denials that affected specific patient groups because cheaper treatments were favored which sparked an ethical controversy about these systems (ComplexDiscovery Staff, 2024).

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 refers to biases that arise from 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.

Finally, 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 stems from a feedback loop when clinicians accept AI recommendations even if they are incorrect, leaving 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.

Impact of Bias

The impact of bias is wide-reaching. Barton et al. (2023) add to the discourse surrounding aided intelligence in healthcare by pinpointing the significant problem of racial equality and machine-learning bias in ML models. Their study highlights the sensitivity of biased healthcare

ML models, especially regarding racial disparity, which is a major issue in disparities compensation and thus is crucial to address.

Barton et al. (2023) seek to answer the question of whether, guided by consideration of racial equity and accountability, healthcare professionals would entrust their patients to AI for diagnosis and disease treatment. This exemplifies a more systematic perspective toward how ethics should be seen as an essential pillar of AI decision-making in the healthcare industry. Furthermore, Celi et al. (2022) have made a substantial contribution to the debate on augmented intelligence in healthcare through an extended assessment of global healthcare disparities and sources of bias created by AI. They demonstrate the excessive bias that exists throughout the AI development process (Celi et al., 2022) by focusing on the health sector, ranging from issues with equitable access to accurate diagnoses and medical advice. Celi et al. (2023) pinpoint the bias inferred from data acquisition, algorithm design, and model training; they also illustrate how healthcare interventions supported by artificial intelligence reproduce systemic patterns of inequality. Bias in the data collection process, including a lack of specific demographic group characters or misrepresentation, is capable of creating biased datasets, which can intensify existing disparities (Celi et al., 2022). Also, algorithm biases in design and model training can extend the existing gap between oligarchs and subjugated by consolidating the prevailing spine of stereotypes and discrimination.

Through their discussion of how the risk of bias in AI may negatively impact healthcare delivery, Celi et al. (2022) demonstrates the critical importance of preemptive strategies to counter this tendency and engender equity. Combating discriminatory behavior and holding each other responsible—through the collaborative work of the main players—can lead to the fullest utilization of AI technologies and enhance healthcare access, equality, and quality of life for all, no matter their race. The theme of AI liability is consistent with the widespread discussion of ethics in AI, where the elements of fairness, transparency, and patient consent are vital.

Ethical Components

Marques (2021) covers the role of bias, information, and interpretation issues, as well as the minority group involved, to counter algorithmic biases in human decision-making (Marques, 2021). Organizations with increased awareness of biases and diversity can address ethical challenges and promote more equitable healthcare outcomes (Marques, 2021), which in the end improves healthcare.

The Ethical Code of Practice concerns the ethical values, standards, and rules that guide the actions of medical personnel and patients (Char et al., 2020). As in the past, ethical issues involve honoring the autonomy principle, providing benefit, refraining from harm, practicing justice, and understanding integrity, and in today's case, understanding how to use AI among others. Ethical issues in healthcare shape various aspects of the duties—for example, patient care, research, resource allocation, and even policy-making procedures (Char et al., 2020). By incorporating ethics into the manufacturing and the development of AI solutions in healthcare, it will be possible to protect against the undesired consequences and risks associated with the use of AI in healthcare; hence, the likelihood of errors will be minimized (Char et al., 2020). The relevance of ethics highlights the need for social and behavioral aspects such as fairness, accountability, and privacy to be included in the design and implementation of artificial intelligence in medicine.

Fairness

Chen et al. (2023) make a remarkable contribution to the ethical discourse on augmented intelligence in healthcare by exploring the concept of algorithmic fairness and its repercussions for AI systems in medicine and healthcare. By scrutinizing the range of fairness-focused techniques in training and post-processing, the authors demonstrate the responsibility of AI designers to construct tools that reduce biases and guarantee equality within healthcare services (Chen et al., 2023).

Applying some of the fairness components to the development or employment of AI technologies can help stakeholders detect and solve problems of algorithmic bias, consequently building and growing trust around AI-assisted healthcare decision-making systems. For instance, fairness-aware training methods consist of customizing the training process to place equity in the center. In after-processing, triggers are used to modify the algorithmic output and reduce bias (Chen et al., 2023). Adopting fairness components into clinical practices serves as a manifestation of ethical AI that believes in the principles of patient welfare and no discrimination.

Further, consistent checks and refinement of AI algorithms by stakeholders will guarantee fairness and equity. With that accomplished, there can be a culture of respect and openness in AI-based healthcare (Chen et al., 2023). By involving data scientists, healthcare professionals, ethicists, and policymakers in the work, the echelons will be responsible for ensuring the development of AI systems that enhance health outcomes, observe ethical values, and improve health equity. Barton et al. (2020), recommend that regulatory bodies should not only minimize bias but also promote fairness in AI-produced health decision-making processes. They encourage AI users to be transparent and take up the responsibility to tackle such issues as the process of developing and deploying ML models.

Accountability: Regulatory Oversight and Governance

Accountability stands out as a critical part of ethical and responsible aspect of augmented intelligence and machine learning use in healthcare. According to Tamvada (2020), accountability describes the obligation of people or institutions to their actions, decisions, and outcomes. In the case of healthcare delivery, increased accountability furthers the preservation of ethical principles and ensures that AI technology advances equitable healthcare delivery. The researchers' discovery emphasizes the need for strict governance mechanisms along with regulatory oversight to manage biases in ML models, combat the unintended consequences, and create trust in AI-integrated healthcare systems.

As pointed out by Larson et al. (2021), regulations involve legal or formal standards—which are usually in the form of the laws and regulations that govern the health sector. Regulatory infrastructures for AI should secure patients' privacy and show ethical conduct by setting the bounds for data security, algorithm validation, clinical validation, and approval procedures (Larson et al., 2021). Health organizations should primarily be proactive. Health organizations must take proactive steps to address difficulties while they benefit from healthcare operations achieving better cost-efficiency and equity. The continuous need to address AI algorithm processes emerges from closely examining these systems and the associated work to remove biases that diminish their usefulness and fairness. To establish accountability, we need to create interpretive paradigms and methodologies that enable analyzing AI systems and assessing their decision-making effects. Ahmed et al. (2020) undertook an exhaustive study of the AI platforms that can be used for healthcare and precision medicine. Their research is a reminder to the healthcare sector that ethics and accountability should be the cornerstone of AI development, considering that responsible use of these technologies is the basis for optimistic health outcomes and patient confidence. Through joint efforts between technologists, healthcare professionals, ethicists, and policymakers, stakeholders argued for robust accountability frameworks that protect against possible ethical hiccups. Through the support of accountability processes within AI platforms, such as transparency of algorithmic decision-making and strong data governance frameworks, stakeholders are equipped with tools for the fair and responsible use of AI in healthcare. This corresponds to an overall enterprise–strategic objective of keeping the AI process open to accountability, including data acquisition, model training, deployment, and evaluation.

Privacy and Informed Consent

Ethical issues play a crucial role in healthcare decision-making, especially when it comes to AI/ML integration. In their work, Wang et al. (2023) shed light on the ethical issues related to ChatGPT, a code that supports communication between people and computers, in clinical environments (Wang et al., 2023). Through an in-depth exploration of how these technologies might impact patient privacy, consent, and even the peculiar tendency of AI-based responses to show a bias, the authors elevate the ethical implications of using these technologies.

The use of conversational AI models such as ChatGPT for medical matters may raise frustrations relating to security concerns and confidentiality of patient information. Such systems usually require access to sensitive medical data to give the right answers (Wang et al., 2023). In addition, customers' informed consent becomes crucial, as they might not fully witness the repercussions of dealing with AI-powered chatbots. Furthermore, any AI system inclination toward biased or misleading outcomes introduces ethical problems, such as untrue or discriminating information that might adversely affect patient care.

Wang et al.'s review of relevant ethical issues urges us to design appropriate governance structures, as well as ethical codes, for AI use in healthcare systems (2023). Questions around AI governance should be addressed by stakeholders so all ethical risks can be managed properly and so that the intelligent model of conversational AI in healthcare complies with principles of autonomy, confidentiality, and fairness. Similarly, the review of AI systems should be continually monitored and assessed to detect and address ethical issues when they arise, which is a facet of being transparent and holding people accountable (Wang et al., 2023). Having a governance framework in place allows critical actors to gain the advantages of AI: having better healthcare service while not violating patient interests at the same time. This research emphasizes the necessity of ethical standards and laws controlling the use of AI in medical decision-making such that individual welfare and autonomy are a key consideration.

Multidisciplinary Approaches

While the data science community has made giant strides in AI-driven healthcare, more so than ever the ethical issues at play and the research in the field prove the need for interdisciplinary collaborations between healthcare professionals, ethicists, data scientists, and policymakers to develop and implement methodologies that mitigate bias and guarantee fairness (Celi et al., 2022). AI development must focus on removing biases throughout its entire process which includes both idea generation and implementation to make healthcare services more accessible and of higher quality for everyone regardless of their socioeconomic background. Aquino et al. (2023) emphasize integrating multidisciplinary perspectives, including clinicians, ethicists, and policymakers, to systematically address epistemic biases and improve fairness and effectiveness in healthcare AI solutions. Learning from multidisciplinary stakeholders increases the ability to discover and thwart biases leading to ineffective resource distribution and soaring expenses. This is an important factor indicating that interdisciplinary collaboration ought to be observed in the creation and implementation of augmented intelligence techniques to ensure effective expense handling. The joint expertise of the ethics, medicine, and computer science fields will help craft AI systems that are attainable and fair in their healthcare delivery services.

Techno-Ethical Integration

ML-augmented intelligence includes compassionate and practical dimensions, in addition to epistemic ethics. Aquino et al. (2023) offers priceless knowledge on the complex facets of algorithmic bias through operation, theory, and morality. Biases in AI systems may supersede healthcare expenditures and disparities; there is an ethical responsibility to deal with these biases to bring fair and equitable delivery services. The authors acknowledge the ethical requirement of getting rid of the biases in the system (Aquino et al., 2023). However, there are other stakeholders and perspectives that are highly relevant to addressing healthcare bias.

Algorithm bias perpetuation is closely related to social justice. Through their elaboration on the interdependency of technical, knowledge-based, and social aspects, Aquino et al. (2023) support embracing a wider-spectrum line of thinking amid the fight against algorithmic bias in healthcare artificial intelligence. Their study suggests that multidisciplinary teams involving clinicians, data scientists, ethicists, and policymakers are key to finding solutions that eliminate biases and techno-ethical issues.

People-Powered AI

Paulus and Kent (2020) drill down into the alarming matter of clinical prediction algorithms and their ability to reinforce the disparity between the winners and losers in the health community, thus dividing and highlighting the complex dance of augmented intelligence (AI) with equitable healthcare delivery. Their study supports the idea of “technology plus humanity” or “people-powered AI.” This refers to the need to note and eliminate disparities to ensure fair allocation of AI technologies across diverse communities. Acknowledging the practical consequences of AI algorithm biases—such as giving the wrong classifications on diagnosis and treatment guidelines—and the wide societal implications of health inequalities opens discussion about the more complex problems of using AI in healthcare to achieve both cost and equity-level outcomes. Their observations bring to light the need to build a more inclusive and ethically correct AI intelligence, to develop AI to handle discriminatory issues, and to design interventions specifically for marginalized and underserved communities.

By incorporating bias and health disparities into considerations for the reduction of AI costs, Paulus and Kent imply that we can ensure that all individuals, regardless of demographic factors such as race and social status, will be the winners of AI solutions. Working together, sharing resources across expert bodies and various sectors of stakeholders, we will be able to unleash the power of AI to improve the access, quality, and affordability of healthcare for the entire world.

Cost Management and Implementation: Cost-Effectiveness Analysis

An article by Paulus and Kent (2020) points to algorithmic health outcome predictions that make some people appear sicker than others, which is rather objectionable in terms of healthcare cost. Thus, without proactive regulatory frameworks to stave off biases, predictive analysis could both worsen healthcare access disparities and lead to higher healthcare costs (Larson et al., 2021). Within the arena of augmented intelligence in healthcare, there are formidable practical matters of implementation and cost management along with questions of ethics.

Investment in artificial intelligence technology has resulted in positive outcomes for patients and treatment systems, but organizations can be more effective in terms of the cost of augmented intelligence programs. This may help overcome the economic question of a healthcare intervention when it comes to how much one gains in a desired outcome and the amount of money spent (Avanceña & Prosser, 2021). By deciding on the value of health service interventions based on their contribution to enhancing health outcomes, improving quality of life, and minimizing resource utilization, cost-effectiveness analysis assumes a mathematical equivalence between the costs and benefits of several healthcare interventions and helps form profitable decisions and resource distribution in healthcare settings (Avanceña & Prosser, 2021). Another benefit is providing the ability to reveal problems regarding AI acceptance in the organization without empirical verification, not to mention resolution of the barriers and facilitators for top-down implementation. The inclusion of qualitative research methods in AI implementation procedures of healthcare institutions will provide professionals with a tool to use for assessing the delicate landscape of the fiscal part of AI implementation when it comes to ensuring transparency and traceability.

Patient-Centered Design

Alongside technical factors, accountability in AI comprises recognition of the perspectives and lives of end-users, like patients, nurses, and doctors. Rines et al. (2022), who conducted a study on psychosocial relationships with food formed by young adults with inflammatory bowel disease (IBD), add much to the healthcare AI discourse. Their articles emphasize the essential nature of moving patient insights into the creation and assessment of AI-based models for healthcare. Their research focused on the stories and lives of those living with

IBD. Because of this, they show the detailed psychosocial elements that influence the process of decision-making and behavior when it comes to health.

AI-based programs shaped upon the needs and preferences of patients not only increase the performance of interrelated technology but also deepen the empathy and comprehension of medicine providers and technologists (Rines et al., 2022). Patient perceptions being integrated into the design phase of AI-powered health systems may affect a positive outcome: allowing stakeholders to provide services that meet the expectations of patients. In addition, monitoring patients whose treatment involves AI technologies to improve the technology and ensure clinical efficacy is an essential part of this engagement effort. Rines et al.'s research confirms the breakthrough potential of patient-centered research, which derives care progress through the introduction of AI-driven technologies, henceforth leading to personalized, inclusive, and equitable healthcare delivery (2022). With ongoing cooperation among the patients, health workers, and technology developers, all stakeholders can together make use of patient perspectives to lead innovation and outcomes in the era of augmented intelligence in the healthcare sector. The current study is aligned with these goals, having been carried out with the aim of better understanding and amplifying healthcare leaders' perspectives.

Recommendations and Applications

In this study, bias mitigation involves identifying, acknowledging, and reducing biases present in algorithms or decision-making processes. Bias mitigation can include techniques such as data preprocessing, algorithmic adjustments, diversity in data collection, and ongoing monitoring to ensure fair and equitable outcomes. It can also include the implementation of human-centered design.

Both human-centered design (HCD) and patient-centered design operate under common principles that prioritize user needs while aiming for user satisfaction. Smith & Nizza (2022) describe the scope of HCD as including user understanding and environmental analysis alongside user involvement across development stages and user-focused assessment and iterative design implementation for complete user experience coverage and interdisciplinary integration. Patient-centered design within healthcare creates customized services for patients which includes their direct participation throughout the design process. The patient-centered approach focuses on empathy and effective communication to build collaborative relationships with healthcare providers which results in systems and services that match real patient needs instead of assumed requirements.

Interpretative Phenomenological Analysis

Pathak, Jena, and Kalra noted that qualitative research methods are not always used for navigating the challenges that arise from the role of technology in social movements (Pathak et al., 2013). Using qualitative techniques, however, healthcare organizations may comprehend the particulars of procedures, perspectives of stakeholders, and culture that influence the suitability of AI integration at all levels. This level of insight allows them to pinpoint possible economies in terms of expenses and find ideal allocation resources for AI implementation.

Chapman and Smith (2002) develop the conversation on augmented intelligence by considering the applicability of interpretative phenomenological analysis (IPA) in unveiling hidden biases and unintended changes from the boosted intelligence algorithms, especially during the dawn of the new genetics (Chapman & Smith, 2002). The fact that they used IPA to assess the sophisticated qualitative issues related to the interaction between AI technologies and society is a demonstration of the need for rigorous qualitative methodologies to unveil complex social dynamics. Through the use of the IPA method, the researcher was able to discover the real-life experiences and perspectives of a person or a community that must deal with humanoid technology-based decision-making processes, so that the biases and moral dilemmas that may also be hidden behind abstract data trends could be explored (Chapman & Smith, 2002). Such an approach allows the researchers to understand intricacies and nuances that could otherwise go unnoticed, affecting accountability in AI-driven decision-making.

By incorporating qualitative methods such as IPA into the evaluation and fostering of AI algorithms, the stakeholders would be in a position to receive a more complex and inclusive outlook of the ethical aspects of AI (Chapman & Smith, 2002). Attention to detail and well-targeted interventions to reduce bias and enhance fairness will follow suit. Chapman and Smith point out the necessity of teamwork that merges social scientists, ethicists, and technologists to approach the ethical challenges linked to augmented intelligence. These cooperative initiatives allow for the creation of a movement toward accountability and responsibility in AI development and operation.

Conclusion

AI bias in healthcare is a salient concern that holds significant ramifications. Bias has the potential to insidiously infiltrate AI systems, resulting in unfair outcomes particularly impacting individuals within vulnerable or underserved demographics. The pivotal role of healthcare leadership in tackling and mitigating algorithmic bias within healthcare settings cannot be overstated. With a clear understanding of their responsibility in addressing this issue, leaders can ensure that every stage of healthcare algorithm development is equipped with the necessary policies and personnel to prevent bias effectively. This research has expounded the theoretical underpinnings of various sources of algorithmic bias in healthcare AI tools and systems, the ethical ramifications of such bias, and practical strategies for bias mitigation that fall within the purview of healthcare leadership.

Reviewing the literature underscores the importance of enhancing the interpretability and transparency of ML and AI-powered systems to address trust concerns regarding unfair machine learning algorithms in healthcare. Increased transparency not only fosters healthcare practitioners' willingness to utilize these decision-making systems but also encourages patients to share critical and confidential data, thus facilitating fairer healthcare outcomes. Moreover, enhanced interpretability grants healthcare officials the social license to collect, use, and share healthcare data accurately, reflecting the epidemiology of specific groups and stakeholders. Leaders must designate specific individuals accountable for the application of algorithmic decision-making tools, ensuring diversity in their backgrounds, education, and gender to mirror the development team. Additionally, fostering a diverse and inclusive team that represents all pertinent groups ensures that training data reflects the true epidemiology of specific populations and allows for continuous cross-checking to eliminate unconscious biases.

Furthermore, the deployment of algorithmic bias monitoring, identification models, and internal system auditors is crucial for ensuring internal accountability and facilitating the prompt identification, reporting, and mitigation of algorithmic biases. Demonstrating trustworthiness by safeguarding patient data confidentiality and security is essential for earning the trust of data providers, thereby ensuring the provision of quality and accurate data.

A systematic review by Haider et al. (2024) reveals racial disparities caused by AI systems in multiple medical fields while stressing the requirement for ethical frameworks and strong regulations to correct these injustices. Research indicates that biased algorithms trained on inadequate data, particularly excluding crucial statistics from racial minorities, can lead to erroneous diagnoses and prognoses, perpetuating disparate and adverse impacts on minority populations (Haider et al., 2024). Algorithmic bias mitigating policies should permeate every stage of the AI and ML tools' life cycle, ensuring adequate sample sizes and comprehensive coverage of factual data about specific groups. Additionally, all team members involved in algorithm development and validation should adhere to medical algorithm checklists to measure fairness, universality, usability, traceability, explainability, and robustness, thereby promoting justice and eliminating potential bias in healthcare AI and ML tools (Medical Algorithms, n.d.).

In summary, addressing algorithmic bias in healthcare demands a multifaceted approach encompassing transparent and accountable leadership, diverse and inclusive team structures, rigorous data privacy measures, and comprehensive workforce training. Failure to take proactive measures against algorithmic bias risks perpetuating systemic discrimination and compromising healthcare outcomes, underscoring the urgency of implementing robust strategies to foster fairness, transparency, and accountability in AI-driven healthcare systems. Consequently, while the challenge looms large, adopting a strategic framework offers a pathway toward achieving equitable healthcare outcomes for all stakeholders.

Chapter 3
Methodology

Research Design

The proposed qualitative research design will use an interpretative phenomenological analysis (IPA) approach. IPA is a methodology crafted to grasp individuals' lived encounters and their process of interpreting them within the framework of their personal and societal spheres (Smith & Nizza, 2022). The research aims to explore the lived experiences of healthcare leaders concerning algorithmic bias in healthcare decision-making and elucidate strategies for promoting fairness and accountability in AI-driven healthcare systems. Employing an interpretative phenomenological approach will allow for an in-depth exploration of individuals' subjective experiences and the meanings they attach to them within the context of algorithmic bias and promoting ethical considerations in AI-driven systems.

The research design will involve semi-structured interviews with healthcare leaders who are actively involved in decision-making processes within healthcare systems where AI technologies are utilized. Participants will be selected using purposive sampling to ensure representation across various healthcare settings, including hospitals, clinics, and public health organizations. Criteria for selection will include leadership roles with direct involvement in the data collection, analysis, or utilization processes.

Data collection will prioritize rich, detailed narratives that capture the complexities of participants' experiences. Interviews will be audio-recorded and transcribed verbatim to ensure accuracy during analysis. To maintain confidentiality, participants will be assigned pseudonyms.

Data analysis will follow an interpretative phenomenological analysis (IPA) approach, focusing on identifying patterns, themes, and meanings embedded within participants' narratives. The analysis process will involve iterative coding, categorization, and interpretation to develop a comprehensive understanding of the phenomenon under investigation. Additionally, member checking will be employed to validate findings and ensure credibility and trustworthiness of the study.

The study will address the central research question and research sub-questions iteratively throughout data collection and analysis. Specifically, the sub-questions will guide the exploration of participants' perceptions of algorithmic bias, strategies for bias mitigation, and challenges related to ethical considerations in addressing algorithmic bias in healthcare systems.

Ethical considerations will be paramount throughout the research process. Informed consent will be obtained from all participants, and they will be assured of their anonymity and confidentiality. The current study adheres to ethical guidelines outlined by the Marywood Institutional Review Board (IRB).

Ultimately, the findings of this research will contribute to a deeper understanding of healthcare leaders' experiences with algorithmic bias and inform strategies for promoting fairness, transparency, and ethical conduct in AI-driven healthcare systems, thereby enhancing the quality and equity of healthcare delivery.

Research Bias

As I began researching the experiences of healthcare leaders in tackling algorithmic bias within healthcare decision-making, I could not help but acknowledge the influence of my own personal biases in shaping this study. These biases, though often unconscious, can significantly impact the research process and ultimately skew the interpretation of findings.

One such bias to confront is confirmation bias, which risks shaping the interpretation of data in alignment with pre-existing notions. Selection bias poses another challenge, tempting researchers to only engage with viewpoints mirroring their own, potentially overlooking valuable insights. Cultural bias, influenced by one's own cultural background, can subtly impact the interpretation of experiences from diverse cultural perspectives, emphasizing the need for diversity in participants. Furthermore, technological bias and professional bias can skew focus toward technical aspects and industry-driven perspectives, respectively, potentially overshadowing broader social, ethical, and organizational dimensions of the issue.

To mitigate these biases, practicing reflexivity, maintaining self-awareness, seeking diverse perspectives, and employing rigorous research methods are essential strategies. Transparently documenting biases and mitigation efforts enhances the credibility and validity of research outcomes, ensuring a more comprehensive understanding of algorithmic bias in healthcare.

My Potential Biases Explained

Admittedly, as someone who has worked and studied in the healthcare technology field for over a decade, I came into this research with preconceived notions about algorithmic bias and AI-driven healthcare systems that can be categorized as confirmation bias. These beliefs could inadvertently lead me to interpret data in a way that aligns with my existing perspectives, rather than remaining open to alternative viewpoints.

There is also the potential for selection bias to creep in. As I identify and engage with healthcare leaders for this study, I must be vigilant against the temptation to include only those whose views mirror my own. This could lead to overlooking valuable insights from individuals with differing perspectives on algorithmic bias and AI-driven healthcare systems. Dr. Maya Angelou once said, “We all should know that diversity makes for a rich tapestry, and we must understand that all the threads of the tapestry are equal in value no matter what their color” (Angelou, 2014). Cultural bias is another aspect that warrants consideration. My own cultural background might unconsciously influence how I interpret the experiences of healthcare leaders from different cultural backgrounds. It is essential to prioritize diversity in participants and remain mindful of the diverse cultural contexts that shape their perspectives.

Given my background and interest in technology, there is a risk of technological bias skewing my focus toward the technical aspects of algorithmic bias, potentially overshadowing the broader social, ethical, and organizational dimensions of the issue. Furthermore, my professional ties to the healthcare industry could introduce professional bias. My experiences within the field may inadvertently shape the framing of research questions or the interpretation of findings, favoring certain approaches over others.

Additionally, advocacy bias is something I need to confront head-on. While I am passionate about promoting fairness and transparency in AI-driven healthcare systems, I must ensure that this advocacy does not compromise the neutrality and objectivity of my research. Demographic characteristics also have the potential to influence my perspective. As a researcher, it is crucial for me to recognize how my gender, race, and socioeconomic background may shape my understanding of algorithmic bias and healthcare decision-making.

To mitigate the impact of these biases, I must practice reflexivity and maintain self-awareness throughout the research process. By critically examining my assumptions and biases, seeking diverse perspectives, and employing rigorous research methods, I have striven to ensure the credibility and validity of my findings. Transparently documenting potential biases and the steps taken to address them further enhances the trustworthiness of the research outcomes.

Sampling

This qualitative study involved 12 healthcare leaders actively engaged in decision-making processes within various healthcare settings. The sampling approach was non-random purposeful sampling, targeting individuals with direct involvement in data collection, analysis, or utilization processes within AI-driven healthcare systems. Additionally, a snowball sampling technique was utilized to expand the participant pool, leveraging the networks and referrals of initial participants to identify additional relevant individuals. Sampling continued until construct saturation took place.

The selected participants represent diverse perspectives and experiences across different facets of healthcare leadership, including but not limited to hospital administrators, clinical directors, public health officials, and policymakers. This deliberate sampling strategy was intended to ensure the richness and depth of data by capturing a wide range of insights and perspectives related to the exploration of algorithmic bias in healthcare decision-making.

Inclusion Criteria

Participants were allowed to participate if they were healthcare leaders actively engaged in decision-making processes within various healthcare settings, including hospitals, clinics, and healthcare technology companies. Several key delimitations delineate the inclusion criteria for this study:

1. The geographic boundaries, focusing primarily on healthcare leaders' perspectives along the East Coast, are vital to this study, ensuring its relevance and specificity to the local healthcare context.

2. The study targets healthcare leaders and executives directly involved in decision-making processes related to data within healthcare settings. This demographic selection enables a focused inquiry into the perspectives of those most closely engaged with data management.
3. The study's population constraints ensure that leaders within healthcare entities of all sizes, from healthcare and pharmaceutical companies to small clinics to large hospitals, that regularly interact with data are included. This approach emphasizes a qualitative approach to capturing nuanced insights from diverse healthcare leaders.

Exclusion Criteria

Individuals were not allowed to participate if they were not actively engaged in decision-making processes within various healthcare settings. Other exclusion criteria that posed a challenge included leaders who met the qualification but spoke a different language and those who could not meet due to scheduling conflicts.

Recruitment Strategy

The study included a sample of 12 participants from the target population. Participants for the semi-structured interviews were recruited through word-of-mouth referrals and contacting provider networks, hospitals, and other relevant sources. These interviews were conducted with individuals located along the East Coast of the United States—in Maryland and Massachusetts to be specific. To protect their privacy, interviewees were told their identities would not be revealed in the write-up of the research and any quotation used would be anonymous. Before each interview, the researcher screened potential participants for eligibility and ensured that they provided informed consent (refer to Appendix A).

Instrument

Interpretative phenomenological approach (IPA) qualitative research design was used to understand how leaders personally make sense of their experiences related to algorithmic bias and decision-making. This technique was used for data collection because it allowed flexibility when focusing on the lived experiences of healthcare leaders by exploring each of their subjective perspectives and interpretations.

The data collection involved interviewing individuals (see Appendix B for a list of interview questions) to understand their subjective experiences, perceptions, and interpretations of the phenomenon under study. Such research studies typically include semi-structured interviews. These interviews are designed to be open-ended, allowing participants to express their experiences, thoughts, and emotions freely. I worked to facilitate a dialogue that encourages participants to reflect deeply on their experiences with algorithmic bias and to articulate their strategies for addressing it.

There are several advantages of using this interview style for the current research. First, having rich detailed data through interviews allow for a deep exploration of participants' insights, providing rich qualitative data. Another positive is flexibility. Semi-structured interviews offer flexibility, allowing researchers to probe deeper into interesting topics or experiences that arise during a conversation. This allows for the individual to not only address the question being asked but to also add their perspective to the conversation.

There are also potential disadvantages to the interview method. For one, interview data can be influenced by the biases and perspectives of both the interviewer and the participant, potentially impacting the validity of the findings. Conducting interviews can also be time-consuming and resource-intensive, particularly when analyzing and interpreting the data.

Additionally, the findings from qualitative interviews may not be generalizable to a larger population, as they reflect the experiences and perspectives of a specific group of participants. Despite these limitations, interviews remain a valuable instrument for collecting data in interpretative phenomenological research, offering unique insights.

Procedures

Once an interested individual is screened as eligible to participate in the study and agreed to do so, the researcher arranged a semi-structured, one-on-one interview, either via phone call, face-to-face or virtual via Microsoft Teams or Zoom. Interviews were conducted at a date and time selected by the participant to ensure the participant could give full, unhurried responses to the interview questions without pressure to attend to other obligations.

The interviews were transcribed verbatim and analyzed using thematic analysis by the researcher. Atlas.ti 25 software was employed to categorize recurring thematic codes from the interviews, allowing for the identification of the most significant issues for participants. Through this process, topics and themes that emerged consistently across interviews were pinpointed, forming the foundation for the analysis and conclusions presented in this paper.

Data Analysis

In this qualitative research, several steps were taken to ensure the trustworthiness of the interpretation of the transcripts, adhering to the procedures for internal validity. First and foremost, a commitment to prolonged engagement with the participants was prioritized, allowing ample time to observe and interact to mitigate potential distortions. Additionally, persistent observation was employed to inquire into participants' experiences, capturing rich and thick descriptions. To enhance the accuracy and reliability of the data, some interviews were recorded, facilitating comparisons with notes taken. Furthermore, self-reflection was embraced to acknowledge and address any biases the researcher may bring to the study. Triangulation was employed by checking multiple sources of data, including input from other interviews, literature reviews, and meticulous field notes. These rigorous methods collectively aim to ensure the true value and internal validity of the research findings, fostering a robust and credible interpretation of the collected transcripts and themes.

Ethical Assurances

The study proposal was reviewed and approved by the university's IRB before the recruitment of participants began. The researcher acquainted participants with the terms of informed consent during initial phone contact, after which participants signed a copy of the informed consent form before the data collection. Participation in the study was entirely voluntary, and participants were informed that they may withdraw at any time, or refuse to answer any question, for any reason, with or without informing the researcher of the reason. The decision to withdraw from the study did not entail any negative consequences. At the same time, there were no incentives for participating. Participants' identities were kept confidential. Names and all potentially identifying details were omitted from interview transcripts, and recorded interviews were stored only on a password-protected folder on Google Drive.

The study followed the principles of the Belmont Report, which outlines several key principles, including maintaining respect for participants, beneficence, and justice. The researcher treated each participant with the utmost dignity and respect. The researcher treated each participant with courtesy, answering their questions honestly, and followed through on all promises made to participants, such as keeping their data confidential and representing their experiences accurately. The researcher ensured beneficence by treating all participants equally. All participants were asked the same protocol questions, and their data and confidentiality were handled in the same way. Finally, the researcher ensured justice, accomplished by sharing the results of the research with each participant and the academic community at large (U.S. Department of Health and Human Services, 2021).

The ethical assurance embedded within this work on AI and healthcare decision-making is foundational to the integrity and credibility of the research. The adherence to rigorous ethical standards not only safeguards the rights and well-being of the participants but also upholds the principles of integrity and accountability within the academic community.

From its inception, the study has been subject to scrutiny and approval by the university's Institutional Review Board (IRB), underscoring a commitment to ethical oversight and compliance with established guidelines. Prior to any engagement with participants, the researcher diligently ensured informed consent, a cornerstone of ethical research involving human subjects. Through transparent communication and the provision of comprehensive information, participants were empowered to make autonomous decisions regarding their involvement, with the assurance that their consent was voluntary and revocable at any stage without repercussion.

Moreover, the researcher took significant measures to safeguard the confidentiality and anonymity of participants. By omitting identifying details from transcripts and employing secure data storage protocols, the integrity and privacy of participants were prioritized throughout the research process. This commitment to confidentiality not only mitigated potential risks to participants but also fostered an environment of trust and respect, essential for meaningful research outcomes.

In summary, the ethical assurance embedded within this dissertation serves as a testament to the researcher's steadfast commitment to upholding the highest standards of integrity, transparency, and participant welfare. By adhering to established ethical guidelines and principles, the study not only advances knowledge within the field of AI and healthcare decision-making but also exemplifies the ethical responsibilities inherent in human subjects' research.

Chapter 4
Results

Introduction

The purpose of this research is to explore the perceptions, impacts, and strategies related to AI bias in healthcare decision-making processes from the perspectives of healthcare managers. This chapter presents the views of healthcare leaders that participated in qualitative interviews to identify bias in AI, its effects, and possible solutions. The background of this study is founded on the rapidly emerging realization that although AI-based technologies hold a great deal of potential for enhancing both performance and patient-centered care, there is also a significant potential for harm if not properly managed. Studies show that bias in medical AI systems generates poor clinical decisions, which intensify and widen existing health disparities. As discussed by Locke et al., (2023) in their work *Preventing Bias and Inequities in AI-Enabled Health Tools*, AI systems can increase healthcare biases if they are not properly designed and tested. Healthcare organizations carry the critical duty of taking additional measures to eliminate built-in AI biases that lead to fair and equitable patient care.

Demographic

Interviews with 12 healthcare leaders were conducted in the fall of 2024 in order to gather data. The study participants included five males and seven females, providing diverse opinions based on various professional positions. The participants occupied a variety of senior positions, such as Chief Technology Officer, Chief Information Officer, Chief Operating Officer, Medical-Nurse Case Manager, Digital Transformation-Enterprise Architect, Medical Doctor/Physician, Chief Medical Officer, and Chief Audit Officer. These individuals were based in Maryland and Massachusetts and worked in a wide variety of organizational environments, including not-for-profit, for-profit, and the federal government. The participants brought a multicultural ethnic tapestry that included African Americans, Caucasians, Indians, and West Indians. The participants were all highly educated, ranging from bachelor's degree, medical doctors, to PhDs. Table 1 depicts the demographic breakdown of the participants.

Table 1
Participant Demographics

Demographic Category	*Description*	*Percentage*
Gender	Male	5 (42%)
	Female	7 (59%)
Job Titles	Chief Audit Officer Chief Information Officer Chief Medical Officer Chief Operating Officer Chief Technology Officer Digital Transformation-Enterprise Architect Medical Doctor/Physician Medical - Nurse Manager	1 (9%) 1 (9%) 2 (17%) 1 (9%) 1 (9%) 1 (9%) 1 (9%) 3 (25%)
Region	Maryland Massachusetts	4 (34 %) 8 (67%)
Industry	Federal Agency For-profit Not-for-profit	3 (25%) 3 (25%) 6 (50%)
Ethnicity	African American Caucasian Indian West Indian	4 (34%) 4 (34%) 3 (25%) 1 (9%)
Level of Education	Bachelor's Masters Juris Doctorate MD PhD	2 (17%) 5 (42%) 1 (9%) 3 (25%) 1 (9%)

The 12 healthcare leaders offered several perceptions of the issues resulting from biased algorithms, and various approaches to AI-bias mitigation. Thus, the findings of this study contribute to the larger conversation concerning the ethical use of AI in healthcare facilities. The findings were organized around five key themes derived from qualitative data: perceptions about AI bias, effects and outcomes of AI bias, ways to deal with AI bias, problems and prospects of AI bias, ways to inform stakeholders and the public, and potential developments in the future (see Table 2). Each theme represented a diversified aspect of AI bias in healthcare. Together, they present a conceptual framework of how leaders perceive and work through these kinds of circumstances.

Table 2
Themes & Findings Summary

Theme	**Key Findings**
Sociotechnical Nature of AI Bias	AI bias is not just a technical issue but a societal one, reflecting historical and cultural inequalities.
	AI systems are endpoints of the data and algorithms that ground them, making bias inherent in their design.
	Leaders emphasize the need for transparency and fairness in AI systems, though opinions vary on the severity of AI bias.
	Regulatory frameworks and organizational culture significantly shape perceptions of AI bias.
Addressing the Impact of AI Bias on Vulnerable Populations	Biased AI disproportionately harms marginalized groups, exacerbating existing health disparities.
	AI bias can lead to misdiagnoses, inappropriate treatments, and unequal resource allocation.
	Biased AI systems erode trust in healthcare technology and create additional burdens for clinicians.
Strategies for Mitigating AI Bias are Needed	Diverse and representative datasets are critical for reducing bias in AI systems.
	Continuous monitoring and auditing of AI systems are essential to ensure fairness.

	Collaboration among interdisciplinary teams (e.g., health professionals, data scientists, ethicists) is key to addressing bias.
	Training and education on AI bias are necessary for healthcare professionals.
Promoting transparency and stakeholder education is critical to aligning AI accuracy with fairness and patient outcomes.	Ensure patients understand AI decisions and fosters trust through transparent communication.
	Ensures patients understand AI decisions and fosters trust through transparent communication, with the "human in the loop."
	Patients have the right to understand how AI-driven decisions are made; involve community stakeholders and feedback systems to improve AI accountability.
The Future of AI in Healthcare	Standardizing data collection and management practices to reduce algorithmic bias and achieve transparent decision-making in AI systems.
	AI governance structures require traceability features to support transparent decision-making processes.
	Continuous training about AI technologies and bias detection as well as adherence to ethical standards that focus on patient well-being and equality within healthcare AI systems.

Theme 1: AI bias arises from the interplay of societal factors and technological design

Interviews revealed that five out of 12 participants stated that AI bias stems from its sociotechnical nature, i.e., both societal factors and technological design contribute to its impact. Perceptions and understandings of AI bias among healthcare leaders constitute a complex landscape that is informed by personal experiences, organizational culture, and the constantly evolving nature of technology. The healthcare executives interviewed understand AI bias but differ in their assessment of its importance to patient equity and safety.

Some participants stated that they were becoming more familiar with the problems of AI bias and its consequences for patients. Speaking about her experience, one leader stated that, “as we incorporate AI into our work settings, we must remember that bias is not solely a technical problem, but it is a problem of society and, if not solved, could impact the lives of our patients." Of course, sociotechnical inequality is already reflected in AI systems and affects vulnerable populations most of all. This encourages the understanding of AI systems not as universal tools but as endpoints of the data and algorithms that ground them. Participants also pointed out the potential for biases to arise within AI systems. One interviewee stated that "bias can be introduced based on how the model is trained and what datasets are used. The results can hugely vary based on training and datasets." Several views were expressed on the quality and availability of training data for AI algorithms, representing a major concern for healthcare leaders. One participant highlighted that "Data is historic, and with that comes the biases of its time."

One of the most striking arguments that came out of the interviews was that of bias salience. Leaders felt that bringing biases to the forefront in conversations could increase awareness and acceptance of AI technologies. One participant argued, "Bringing biases to the forefront of conversations is crucial. It not only raises awareness but also helps us collectively address the challenges and build trust in AI technologies." Moreover, when addressing human biases, it allows patients to view AI as something that can bring about fairness. However, there was a level of disagreement among participants to the extent to which AI is biased. Some went as far to question the seriousness of the problem, stating that, for example, "AI is just a tool that requires careful oversight to ensure they meet ethical standards, especially when they impact patient data and decisions." This helps in understanding the core risk associated with AI technologies: While some leaders have called for proactive steps to minimize bias, others underestimate its potential impact on clinical decision-making. One interviewee stated that "decisions should not just be made with intelligence, but with empathy; the relentless pursuit of advancement should not overshadow the fundamental need for connection, compassion, and care between the patient and the provider." The allure of profit over purpose and financial pressures exerts a strain on healthcare organizations to earn revenue while not spending much; this pushes the goal of short-term financial gains ahead of the equity of the working agenda.

The establishment of regulatory frameworks has significant influence over public perceptions of AI bias within healthcare systems. One interviewee commented that regulatory standards today focus on swift technological development in AI applications while neglecting sustainable practices and ethical considerations. The pursuit of technological development generates difficulties in bias mitigation because regulatory standards typically neglect essential fairness and transparency measures within AI systems. Countries such as those in the EU are ahead in developing AI healthcare policies, while the US is still catching up, potentially leaving policy decisions to individual states. As one interviewee pointed out:

> Part of the EU initiative on AI regulation—whether it's the EU AI or the AI Act—is the standards organization from the European Union for artificial intelligence and automation. They have, I think, 67 families of controls, each with very specific elements. You could say that they've created libraries of code that will search for bias in AI. They will also search for algorithmic inferences that may not necessarily align with the intended outcome. So, they're creating libraries that can assist with adaptation and help automate some of that work. However, it's still very early in the stages of maturing this.

Regulatory bodies are beginning to address AI's role in healthcare, particularly around utilization management, and concerns over AI-driven auto-denials. Therefore, leaders must push back against overreach in AI regulations while ensuring that AI governance is robust and embedded in organizational processes. Healthcare organizations are in the middle of two processes: technological development and legislative demands. Participant two shared how the FDA's AI and ML software is designed to detect and address biases in healthcare algorithms, ensuring that these technologies provide equitable and accurate outcomes for all patients. In 2021 the FDA created ‘Action Plan for AI/ML-based Software as a Medical Device’ (SaMD). The FDA's action plan is aimed at detecting potential biases in healthcare algorithms (U.S. Food and Drug Administration, 2021) that may results from societal factors interacting with technological design. Healthcare leaders’ knowledge and attitudes regarding AI bias are complex and dependent on these policies, along with factors such as individual experience and organizational culture.

Despite the increasing concern regarding the effects of bias in AI-based systems, discord emerges based on the degree, acceptability and mitigation of the problem. One leader stated that, “We thought we nailed data governance, but the rise of AI shows we need to start fresh, focusing on value-driven outcomes.” This opens a possibility of other studies to further investigate the consequences of AI bias in clinical decision-making.

Theme 2: AI bias harms vulnerable populations and deepens inequities

AI has the potential to alter patient care and the practices of healthcare professionals. However, AI bias has the potential of compounding the adverse discriminations patients are subjected to. A specific challenge was noted by one interviewee:

> AI could push patients away from needed services due to biases in the system, such as when data indicates a primary care provider is nearby, but the recommendation pushes the patient toward more distant options. AI system biases have multiple consequences which prevent patients from accessing suitable healthcare providers located nearby. If training datasets fail to fully represent certain regions or demographics data gaps or errors emerge which lead to AI systems missing nearby healthcare providers. The AI system may overlook the possibility of referring patients to a nearby local clinic. Algorithmic bias emerges when AI systems focus on cost savings by selecting distant providers which align with financial incentives like reimbursement rates instead of local options which might be more convenient for patients. The AI may operate under the belief that providers who are farther away provide better quality services or have greater availability compared to nearby options. The system might choose a distant hospital over local options by using historical data which shows that greater distance usually means higher specialization

> levels. The AI might incorrectly suggest distant options for users because errors in geographic data or outdated maps create misleading data which prevents the system from recognizing closer alternatives. The issues demonstrate that healthcare recommendations become less appropriate when biases exist alongside incomplete data or faulty algorithm design.

This situation exemplifies the tension between data-driven decisions and the need to prioritize patient access and outcomes. Healthcare disparities are also related to the cultural stigma that shapes care-seeking behaviors in African American and Asian populations. For example, one interviewee shared that “biases in algorithms can lead to delayed diagnosis”; their statement was backed with the following comments/example:

> Black men are 50 percent more likely than white men to experience a stroke in their lifetime, but fewer of them seek care due to cultural stigma, a problem that also affects certain Asian communities, like Chinese and Japanese populations. Studies from the AMA [American Medical Association] and Harvard researchers have shown that these groups are less likely to seek care because of these cultural barriers. Simply providing the same treatments to these populations as those given to white individuals, based on biased data, will not lead to the best outcomes. Effective care requires treating people in a way that reflects their cultural and demographic realities for African American men, who are fifty percent more likely to suffer strokes than their white counterparts.

Intelligent technology can be effective in ensuring that patients receive quality and equal treatment when accurate data input occurs. Healthcare leaders report AI bias effects as transcending clinical practice, influencing operational workflow and the dependability of AI systems. Clinical outcome effects demonstrate the greatest misuse of AI. Several interviewees mentioned the fact that biased algorithms could result in misdiagnosis or unsuitable treatment recommendations. One leader stated that, “some of the common areas of bias are socioeconomic bias, racial and gender bias, drug effectiveness. These biases can be addressed using diverse datasets for training and making specific considerations while building and testing the algorithms.”

The performance of AI models tends to be biased, since training data may be inadequate and not representative of real-world scenarios. The accuracy of AI-driven decisions becomes impaired when data fails to represent the target population properly or when it lacks completeness. The examples given above show that if AI systems are not well designed and not supervised, they could further extend disparities in care as exemplified in one interviewee's comments:

> …you should look at this case that's currently unfolding, which has recently been revisited due to a tragedy last week. A company called naviIHEALTH is used by many highly regarded healthcare systems. This company coordinates what many people in the industry would agree is a worthwhile automation, as it can help reduce healthcare costs. For instance, if Amazon had humans manually reviewing all the data to suggest purchases, we'd be frustrated with them. In this case, the company automated a predictive tool to assess which patients would benefit most from staying in a nursing facility or being discharged from the hospital. The tool calculates a risk score, which is then used to determine if a patient can safely be sent home or should stay in a facility. This predictive tool is similar to a risk score used in hospital discharge planning, and many healthcare systems use such tools to direct social workers to the right patients. However, the popular media now describes this as using AI to deny patients access to skilled nursing facilities. I believe the company used this tool in the prior authorization process, which was intended to streamline discharge planning and reduce administrative costs—so more money could be allocated toward actual patient care rather than administration. The way it's being framed now is that machines, rather than humans, are deciding who stays in nursing facilities, which is causing a loss of trust. The lesson here is that automation can only work if there is sufficient trust in the system. People trust companies like Netflix, but they don't always trust healthcare plans to make decisions using AI. We need to be cautious about automating processes, particularly when it involves human lives. The case seems to demonstrate that trust is easily lost, and what could have been seen as a cost-saving and efficient tool is now being perceived as a sinister practice. It's likely that people, well-intentioned though they were, may have tried to speed up processes, but it's painful to see how the system has been described now.

The deployment of AI and automation in healthcare requires trust since errors and misinterpretations could damage public credibility. Automation presents opportunities for increased efficiency but requires careful implementation to avoid adverse effects on patient care and public perception.

Leaders also understood how biased algorithms impose undue burdens on the clinicians who must correct flawed predictions. One interviewee warned that "AI systems, which predict healthcare outcomes based on document data, must identify biased documentation, because such biases could train models inaccurately." They went on to say that the "increased workload from correcting these flaws results in clinician dissatisfaction and obstructs the delivery of quality care." Healthcare organizations also may face trust issues with technological reliance due to the threat of AI bias, according to participants. Patients must understand the reasoning behind their healthcare decisions, because any erosion of trust between them and their clinicians could jeopardize their therapeutic partnership.

> AI is in every footprint, but without understanding its biases, we risk amplifying inequities. As one interviewee state: AI and data governance are rapidly evolving, similar to the challenges faced with enterprise data management in the '90s and 2000s. Despite past efforts to establish effective data governance, we realize that many organizations didn't do as much as they thought. It became a routine, business-driven process rather than one focused on achieving real outcomes or understanding…to reduce disparities and achieve equity, organizations must become data-driven and focus on inclusion. This requires everyone, from data entry personnel to engineers, to understand and play a role in this transformation.

Healthcare organizations generally avoid taking risks which makes it tough to reach goals such as reducing biases in healthcare unless there is complete organizational commitment and understanding. Technology and AI are transforming healthcare operations, but data governance issues persist which requires organizations to adopt data-driven methods that emphasize inclusion and equity to bridge healthcare disparities. Along with this transparency in AI operations need to be placed front and center to enable patients to trust technology and feel empowered instead of anxious about its risk.

The ethical implications surrounding AI bias cannot be overlooked. As health leaders address these issues, they face moral questions relating to fairness and justice. In view of the rapid pace of the adoption of AI technologies, firm commitments to ethical principles are necessary, with the interests of patients first. One participant emphasized this point:

> Healthcare's risk-averse nature makes a risk mitigation framework for ethical AI an essential layer of protection. This is a risk management strategy since healthcare is very risk-averse. You can always win with a conversation about risk, and while it may seem like a scare tactic, if you look at a risk mitigation framework for enhancing and enriching data governance to address bias and ethical AI, it adds another layer of protection for your organization and the members you serve. So, while you can approach it from a "stick" perspective, the "carrot" is that you can do good and do good at scale.

This emphasizes the positive impact that can be derived from efforts that integrate the need for philosophical frameworks that guide the decision-making processes on whether to deploy AI. Participants also expressed concerns regarding the ethical consequences of decision-making biases in artificial intelligence systems. AI advancement proceeds quickly, which requires experts to maintain a balance between innovative technology and ethical guidelines that protect patient welfare. Participants stressed that organizations must create strong ethical guidelines to protect equal healthcare access because innovative needs must reconcile with the ethical responsibilities owed to patients. As one participant remarked, "It is imperative to meet the needs of innovation while considering the ethical duties of patients equally and fairly." Regulatory functions are essential for mitigating AI bias, yet they face difficulties maintaining pace with fast-moving technological progress. The swift advancements of AI technologies create growing difficulties for regulators who must manage and mitigate new forms of bias.

Theme 3: Strategies for mitigating AI bias are needed

Eradicating bias in healthcare applications of artificial intelligence is complex. The healthcare leaders interviewed are seeking ways to manage the problem effectively. From the interviews, it is clear that the use of structure for reducing bias in the AI life cycle is highly appreciated. The use of diverse and representative data is among the most basic strategies for bias reduction in AI. Twenty-five percent of the participants stated that model building and data collection should mirror the population they are serving, however, inclusive data collection practices are one of the intrinsic elements in non-discriminatory AI systems. Healthcare organizations need to constantly refine their data collection and management systems to achieve inclusive data. “Through proactive data collection—which represents various demographic groups including age, gender, ethnicity, and socioeconomic status—algorithms will be able to deliver optimal healthcare service options for patient populations.”

Throughout the different conversations there was an emphasized requirement for strong auditing mechanisms, which would provide routine assessments of AI performance in different groups. One respondent stated that “AI is highly disruptive, and it is still growing; therefore, it is critical to have the correct strategy for adoption.” However, mitigation strategies should be developed considering different sources of bias from varied stages of the AI algorithm's development process. Hence, “setting protocols for continual evaluation acts as a way for healthcare organizations to ensure that their AI systems remain effective and equitable over time.” In addition, “fostering collaboration among interdisciplinary teams to work together in developing and implementing technologies of AI” is key. In the design phase, involving health professionals, data scientists, ethicists, and community representatives allows for a much broader consideration of potential biases and their consequences.

Training and education about AI can go far in helping healthcare professionals learn how to identify biases as well as finding solutions. One participant stated that, “Equity starts with understanding your role in the process, whether you’re entering data, managing claims, or building APIs [application programming interfaces].” An API acts as a messenger that allows different software programs to communicate and share information with each other. Therefore, implicit bias training can enhance the decision-making profile within healthcare facilities.

Ethical considerations must also guide mitigation efforts of AI bias. What the leaders have pinpointed are framed ethical frameworks, underlining fairness in all decision-making processes regarding the use of AI. Ensuring that AI systems not only perform accurately but also with fairness is crucial, especially in healthcare situations where patient outcomes are directly impacted. For trust and accountability for AI applications in healthcare to be achieved, ethical challenges in machine learning need to be faced. However, the dynamic legal requirements might serve as guidance toward defining organizational practices about bias.

Theme 4: Promoting transparency and stakeholder education is critical to aligning AI accuracy with fairness and patient outcomes.

Leaders must strike a balance between transparency requirements and open discussions, which are essential for addressing AI biases in healthcare according to participants. It was clear from several interviews how important trust along with clear and transparent communication is when it comes to making decisions and interacting with patients. Healthcare stakeholders need to engage in effective communication to address AI bias.

Answering the question on how to communicate with patients and the public about AI bias in decision-making, focusing on transparency and trust, several interviewees found it essential to communicate safety and security protocols as AI technologies become integrated into clinical workflows for both providers and patients. One interviewee shared that “healthcare professionals must take proactive steps to engage stakeholders when it comes to reducing AI bias to support equal treatment for all patients.”

Many participants identified transparency in AI decision-making as a significant challenge, however. They emphasized that patients deserve to understand how healthcare decisions are made. One respondent noted that “patients need to understand the decision-making processes because their trust in healthcare diminishes when they perceive bias within the system—and through transparency, healthcare systems will be able to build trust while patients gain the necessary information to participate actively in their treatment decisions.”

Analysis and findings revealed that the context in which these cultures are performed within the healthcare organizations plays a major role in how leaders perceive the bias in AI. Due to the complexity of patient care and the potential for high-stakes consequences; leaders need to educate their staff to ensure AI tools are applied correctly and ethically, ensuring transparency and fairness in decision-making. In healthcare, the choice that leaders make might cost people their lives. While this caution may stifle innovation, it is a virtue of the patient safety movement. One of the participants stated that, “My view is while AI must be adopted in healthcare, it is critical to ensure the required process and organization are put in place while still keeping the agile mindset.” This reflects the tension between the need for strong procedures and ethical AI implementation in healthcare decisions as leaders navigate the balance between inventive approaches and careful advancement.

Participants also highlighted community engagement as an essential tool to identify and understand the specific needs of different population groups. One participant stated, “involving community stakeholders delivers valuable information on AI technology impacts across various groups.” Through community discussions about AI deployment, healthcare systems can build technologies that meet the needs of diverse populations.

Education was also shared as an essential tool to counteract AI bias. Several healthcare leaders emphasized the importance of training workers to identify and correct biases present within AI systems. As one interviewee stated, “it is essential for leaders to provide training to their staff for the deployment of AI systems that operate without bias and maintain fairness.”

Finally, the participants identified effective communication strategies with patients and the public as a crucial element. Participants advocated for patients to become knowledgeable about the technologies that transform their healthcare. Participants identified the need for feedback systems that enable healthcare organizations to communicate bidirectionally with the communities they serve. When organizations encourage patients to provide feedback on AI decisions, they encounter technology’s weaknesses and hidden flaws. Thorough conversations about data-sharing policies and patient rights regulations are necessary to develop AI practices that are both transparent and accountable.

Theme 5: The future of AI in healthcare

Healthcare leaders will need to outline a transparent roadmap when integrating AI into healthcare, with a strong emphasis on ethical considerations and improving patient outcomes. Interviewees believe that AI technology has the potential to revolutionize patient care, but also stress the importance of actively identifying and addressing biases within healthcare systems. Overall, experience has shown most leaders the critical importance of data standardization and data management. Leaders warn that AI models face elevated bias risks when there are no standardized methods for data collection and usage. The following statement from a participant highlighted the critical need for setting clear data collection and usage standards:

> We need to define clear standards of how data should be collected and used; otherwise, we will continue with biased algorithms.…one of the primary frameworks we’re adapting is the AI governance structure, which includes a key principle that data must be

> thoroughly quality-controlled before being input into the AI system. We're committed to ensuring full traceability of AI decisions. This means we want to understand why a specific answer was generated, including identifying the data sources that contributed to that decision. The focus is on maintaining accuracy and transparency, so we can always trace the reasoning behind any AI-generated decision and ensure its foundation is solid and correct.

The interviewee stresses the critical need for transparent and traceable AI decision-making processes within healthcare applications. The AI governance framework implemented by their team permits quality control for data entering AI systems while maintaining traceability for every AI decision back to its original data sources and reasoning methods. The statement emphasizes the importance of establishing clear methods to comprehend AI systems' decision pathways as well as their underlying reasons to maintain accountability and accuracy. Successful AI model training relies on both quality data and proper methods of applying that data to achieve unbiased and accurate results. The system needs to process and interpret data accurately while ensuring its application leads to trustworthy and equitable decision-making. The performance of an AI system relies on the quality of data available, and the training methodology used to produce insights that are meaningful and transparent with sound reasoning.

Summary

Overall, healthcare organizations can mitigate algorithmic bias by implementing frameworks and standards that help AI systems represent its service population's diversity through inclusive data collection, continuous monitoring for fairness, and regular updates to models that account for demographic changes. The development of transparency and accountability within organizational culture holds equal importance. Leaders believe that transparent dialogue about AI capabilities, limitations, and potential biases is crucial to establishing trust between healthcare professionals and patients. One participant succinctly put it, “Transparency is paramount; it should be a top priority because it creates a trustworthy environment that lets patients participate actively in their care decisions when they know that the technology has undergone rigorous fairness evaluation.” Several leaders stressed the significance of cross-discipline collaboration to speed up efforts for minimizing AI bias through sector-wide cooperation.

There is prevailing optimism about AI’s future capability to identify and correct bias. Leaders anticipate the development of more advanced tools and frameworks as AI technology continues to progress. One leader stated that “In the future, I expect more standards defined to handle bias with a standard governance model.” The discussion highlighted the essential requirement for ongoing education and training programs for healthcare professionals. When staff members gain expertise in AI technologies and bias detection they will be better equipped to recognize and address emerging issues effectively. One leader emphasized, “We have to train our staff on AI bias.” This ongoing education is critical for staying ahead of potential biases and ensuring that healthcare systems remain responsive and accountable to their patients’ needs.

All participants highlighted that ethical considerations should guide the creation and execution of AI technologies. The belief that technology should always serve to improve human welfare instead of worsening existing disparities is reflected in the statement made by one leader, who said, “Ethics comes first, before innovations, because we mind our patients’ well-being.” The integration of AI into healthcare systems necessitates the preservation of an ethical framework to guarantee equitable benefits for all patients.

The interviewees see growing AI integration as raising important concerns about bias in decision-making because of its significant effect on patient care and the continued existence of health inequities. The findings from the study depict healthcare leaders’ conceptions of AI bias in decision-making, its significant impact on patient care and inequity in health. Participants regarded AI bias both as a technical and a social issue, one with origins in current and historical social and cultural inequity. Biased algorithms, built through unbalanced training datasets, can result in misdiagnosis, inappropriate treatment, and unequal distribution of resources, with a bias toward marginalized groups. Leaders emphasized diversity in datasets, continuous monitoring, and multidisciplinary collaboration in overcoming bias.

Ethical concerns took center stage, with a call for frameworks prioritizing patient welfare over innovation. Transparency in AI decision-making was a critical factor in maintaining trust, with patients having a right to information regarding how care-related decisions are reached. Nevertheless, budget constraints and financial incentives consistently obstruct effective countermanding of bias and compel organizations toward prioritizing short-term gain over long-term fairness. Healthcare professionals' training and education in bias awareness and reduction were seen to be critical. Community participatory approaches and uniformed practice with data were also discussed as important strategies for developing fair AI systems. Overall, the findings portray the complexity of AI bias in healthcare, with a strong message regarding the imperative for fair, transparent, and participatory approaches in developing AI technology that will work to resolve, not exacerbate, health inequity.

Chapter 5
Discussion

Introduction

Integrating AI and ML into healthcare systems is as game-changing as it is utterly daunting. Influencing patient outcomes, trust, and equity, the issue of algorithmic bias becomes a central factor in how AI-driven tools shape the decisions made in healthcare. This study aimed to understand the lived experiences of healthcare leaders who manage algorithmic bias to explore how bias happens, what impact it has, and what can be done to reduce bias in AI systems. By using interpretative phenomenological analysis (IPA), this research attempted to understand the practice of healthcare leaders in how they perceive and respond to challenges arising from algorithmic biases. This paper articulates how technological innovation and ethical consideration are tied together and calls for healthcare leaders to create a culture of transparency, human oversight, and inclusivity in AI-based decisions. The findings from the IPA are integrated with existing literature in the discussion section, and actionable policy recommendations are proposed to promote ethical AI adoption in healthcare.

This study uncovered several themes relating to perceptions, impacts, and mitigation of AI bias in healthcare. Algorithmic bias was emphasized to be complex, often empowered by historical inequities and systemic disparities within the data used. Nonetheless, qualitative data showed a consensus among healthcare leaders that bias needs to be addressed and understood to assure equitable patient outcomes.

Sociotechnical Nature of AI Bias

Healthcare leaders have become more aware of the technical and systemic issues that AI bias creates. Nearly all participants referenced bias as the “historical baggage” of datasets that reflect the historical roots of inequities in healthcare delivery. One participant discussed how data comes with the biases of its time. Another participant identified the need to acknowledge the biases as both a technology flaw and as a mirror of the social inequities in healthcare practices. These insights align with Cho's findings that bias is present in AI systems in every possible way (2021). Nevertheless, such commonality does not imply it was possible to mitigate bias completely. In some cases, participants were skeptical about neutrality as an ideal end state. Others said the goal should be to do as little harm across as many people as possible. This is consistent with Ahmed et al.'s (2020) demand for a balanced way to mitigate bias while recognizing technical constraints and moral demands.

Addressing the Impact of AI Bias on Vulnerable Populations

AI bias directly impacts patients' clinical outcomes and affects AI technology's credibility in the sector. For example, as reported by participants, zip code-based analytics push patients to providers based on financial incentives instead of proximity or suitability. A healthcare leader dubbed this problem "a failure to balance data optimization with patient-centered care." The interviews also revealed that marginalized communities are disproportionately harmed by AI bias. Findings about unfair effects for Black and Asian populations in particular underscore the critical need to address AI bias and cultural factors in healthcare, as they directly contribute to delayed diagnoses and unequal access to care for vulnerable populations, further exacerbating existing healthcare disparities (Cross, Choma, & Onofrey, 2024).

Strategies for Mitigating AI Bias

Participants outlined several technical, organizational, and cultural bias mitigation strategies. A common call was the inclusion of diverse demographic and behavioral data.

According to one of the leader interviewees, “all these algorithms start with inclusive data, as their inclusive algorithms begin with inclusive data.” This view is further supported by Hanna et al. (2024), who stated that mitigation strategies should be developed considering the different sources of bias originating from various stages of the AI algorithm’s development process. In other words, training inclusive algorithms requires high-quality data that encompasses diverse inclusivity parameters. Ahmed et al. (2020) also indicated that inclusive data collection practices are one of the intrinsic elements in making non-discriminatory AI systems. The foundation of an inclusive and fair algorithm requires data that represents diverse demographic groups with varied behaviors and cultural backgrounds. If the initial data lacks inclusiveness, the algorithm will fail to produce equitable results for all population groups. Achieving inclusivity of data requires systemic collaboration between data scientists, clinicians, and policymakers.

Another key strategy that came out of the interviews was cultural competency training, which helps healthcare professionals recognize and intervene against biases before they occur. To encourage this kind of training, participants recommended that it become integrated into organizational policies, stating that, “Training like this would help us bridge the gap between technology and human-centered care.” Accountability and trust were also emphasized, along with the need for transparent governance. This echoes Vayena et al. (2018), who argue that for trust and accountability in AI applications in healthcare to be achieved, ethical challenges in machine learning need to be faced. Leaders also called for the rollout of “proof of no bias” assessments, used after the deployment of AI models to ensure fairness and equity. This ties into the persistent review and continuous refinement of AI models. The harmful effects of AI bias on vulnerable populations emphasize the need for ongoing supervision and improvement, with shared responsibility in AI–human partnerships used to increase ethical results and prevent further damage to marginalized communities (Cañas, 2022).

Ethical Tensions in AI Adoption

Ethical AI adoption demands addressing biases through careful solutions such as diverse datasets, transparent model evaluation, and quantifiable validation to achieve equitable outcomes across patient groups. Comprehensive regulatory statutes created by the healthcare community will help achieve successful patient protection standards. Modern healthcare AI solutions can enable improved medical results while examining healthcare operations and enhancing decision processes. AI technology achieves lower operational effectiveness levels when biases remain unregulated, thus damaging both patient trust and operational outcomes (AI Factory, 2024).

Healthcare algorithms that contain prejudice influence medical provider diagnostic accuracy and clinical recommendations and produce unjust differences in how care resources are distributed to patients. Because of these issues, healthcare service disparities tend to develop primarily within underprivileged and minority communities. AI diagnostic tools derive their results from input data, so when these tools are trained on datasets that exclude certain populations, their recommendations will disproportionately benefit some demographics at the expense of others. Biased AI systems cause patient harm, weakening public trust in artificial intelligence systems. AI systems require unbiased frameworks that strengthen regulatory measures through bias testing and removal requirements so that these technological solutions achieve both functionality and fairness.

While creating regulations for AI in healthcare is important, ensuring they preserve the technology's capabilities to transform healthcare is essential. Excessive regulations and strict guidelines prevent advanced AI applications in healthcare from reaching their full innovative potential (Cohn, 2024). Although AI can advance patient care and operational effectiveness through diagnostic tools and automated workflows, regulatory overreach remains a barrier to the implementation of these innovative solutions. Healthcare organizations cannot test new AI solutions quickly when strict regulations prevent the rapid implementation of technologies that might deliver major improvements. A fundamental challenge thus exists—balancing the need to create regulations that provide safety assurance and the importance of allowing AI technology advancements to reach their full potential. AI systems that incorporate ethical principles will enhance healthcare delivery while producing advancements that guarantee safe and effective patient care for all individuals.

Potential consequences of biased AI extend far beyond patient treatment; they actively adjust foundational assumptions of healthcare delivery systems. Present-day AI algorithms may utilize generalized patient data for people of color but disregard important cultural distinctions and living environments that impact healthcare results. This lack of detailed understanding could mean that treatment plans developed for such patients are insufficient for their needs. An algorithm may be able to offer a treatment plan and yet fail to consider fundamental factors that define where the patient may live, work, and play. This oversight is detrimental to quality and puts patients and other healthcare providers on the back foot when it comes to AI.

Additionally, the impact of AI bias on resource allocation is another critical concern. When using an AI algorithm for the allocation of resources within the healthcare setting, where Black patients have received systematically poor healthcare, the allocation is determined by healthcare costs. This translates into Black patients paying lower fees in healthcare and the algorithm then learning they needed less, resulting in the population being further underfunded and neglected. This finding serves to underscore the higher need for organizations to consider AI system design and deployment holistically to avoid deepening existing issues.

Balancing AI Accuracy with Fairness and Patient Outcomes

Trustworthy AI systems in healthcare require a regulatory framework that supports innovation while upholding the medical responsibility to achieve better patient outcomes through an equilibrium of accuracy and fairness. An essential regulatory setup should promote openness and responsible design to support testing freedom, which will generate trustworthy yet revolutionary AI technologies. In 2023, the US Executive Order on AI issued guidelines to enhance AI safety and trustworthiness through fairness and security (Jones et al., 2025). As AI technologies continue to advance, federal agencies must give way for state and local jurisdictions to manage regionally tailored and industry-specific needs. AI systems are predicted to influence business decisions by 2029 and will deliver application development automatically by 2027 without human assistance (Gartner, 2024). Healthcare organizations require up-to-date information about legal and standard changes to maintain compliance and adaptability in this realm of AI healthcare service delivery. AI healthcare tool developers must adopt strategies that confirm product safety and equitable access (Kim et al., 2025). The potential of AI can only be harnessed by medical practitioners who establish open systems for AI distribution based on fast tech development and proper ethical principles, which in return strengthens patient outcomes while sustaining public confidence in health services.

For healthcare industries to realize fully enabled AI capabilities, they must incorporate foundational principles of responsible AI usage into structured approaches. AI systems must also align with developmental ethical agreements and standards inspections to influence their clinical achievement measurements. Healthcare system operators must thus adopt AI systems flexible enough that medical experts can respond instantly to clinical complications that may arise from unexpected biased algorithmic behavior.

Algorithmic bias risks require regulation through AI principles. However, we need the right balance in order to operate effectively. When safety-focused regulations match pace with technological advances, artificial intelligence systems perform better, deliver improved outcomes, support medical research, and ensure operational effectiveness. Healthcare AI systems developers require regulatory mechanisms that provide adaptive and transparent tools while managing ethical requirements. Medical professionals thus need to push for improved regulatory guidelines to ensure that AI advancements meet operational safety controls to achieve patient treatment fairness and maintain patient trust.

Resource Constraints and Equity Trade-offs

Healthcare leaders must find ways to manage resource constraints while ensuring that AI technology reaches all populations fairly. The leaders shape the ethical parameters of AI systems integration within healthcare facilities. Healthcare institutions must develop a workplace that promotes inclusion, complete transparency, and accountability. As they establish ethical problem identification, expert leaders should simultaneously support innovative AI technology implementation. The battle to create systemic change in healthcare requires field leaders to champion the modification of existing healthcare policies to address inequalities found in healthcare data practices. This process requires advocating regulatory systems that hold organizations responsible while allowing every person to experience unbiased advantages from

AI technologies. Additionally, leaders must dedicate their energy to effectively educating and training their staff to understand what AI bias entails.

Additionally, healthcare leaders should incorporate augmented intelligence when merging human expertise with AI skills. One interviewee said, "An AI system should not be replacing human judgment, but should complement it by using AI technology to give clinicians tools to help them make better decisions." Accordingly, this aligns with the augmented intelligence literature and Hacker's argument for a balance between efficient technology and a moral output from the system (Hacker et al., 2023).

This study shows AI's potential in healthcare, but ethical problems of algorithmic bias present barriers. Securing healthcare leadership as a nexus to addressing these problems will require leaders to adopt an environment of equity, transparency, and accountability. Suppose an algorithm was truly inclusive and adaptable. The associated cultural competency investment in building algorithmic literacy and transparent governance would facilitate the adoption of ethical AI technologies aimed at patient-centered care. However, I cannot overstress the necessity for further research and cooperation in a dynamic healthcare landscape. Algorithmic bias demands a long-term solution, and thus, continuous efforts, vigilance, adaptability, and fairness are required. Therefore, these results provide inputs necessary for the development of more general ethical AI for healthcare and lay a foundation for future work. The success of AI-driven healthcare depends on transparent and inclusive leadership, alongside strategic resource management to ensure universal benefits and avoid increasing current disparities.

Recommendations

The development of inclusive algorithms—alongside cultural competency training and transparent governance frameworks—must become our top priority to achieve responsible and fair artificial intelligence integration in healthcare. The establishment of a multidisciplinary partnership between medical experts, data scientists, ethicists, and leaders enables us to minimize biases to the greatest extent possible while creating AI solutions that provide equitable services for all groups. The adoption of ethical AI governance alongside improved data transparency will build trust and accountability that ensures AI technologies improve healthcare outcomes instead of damaging them. Healthcare leaders need to embrace these core strategies and maintain ongoing inclusive and adaptable methods in AI development to achieve true integration between humanity and technology.

Developing Inclusive Algorithms (DIA)

Artificially stored data have wide implications in healthcare, where the ideal algorithm is inclusive, and the accuracy of the results is highly prioritized. Just as with humans, if AI algorithms are not trained appropriately, they can add to systemic disparities. Training datasets require diversity but become predatory when marginalized populations are sampled without guaranteeing fair representation and benefit. Without actual stakeholder status in AI development, marginalized groups remain mere statistical entries. This perspective permits data exploitation and prevents real healthcare improvements. Consequently, models utilizing machine learning with diverse demographic, cultural, and behavioral nuances are crucial. Inclusive algorithms make equity necessary by enabling systems to process and analyze data in a way that does not reinstate past or current inequalities.

Collaboration between healthcare professionals, data scientists, and ethicists is crucial to overcoming algorithmic bias. According to Ueda et al. (2024), fairness-awareness methods are necessary to ensure equity is distributed throughout the machine-learning pipeline. Healthcare leaders should customize training processes to scale underrepresented populations and adopt post-process techniques to mitigate algorithmic outputs. That way, algorithms will not marginalize a group inadvertently. Also, organizations must commit resources to tools that audit and validate algorithms post-deployment. Interview participants revealed that implementing every healthcare algorithm requires standard proof-of-bias-elimination. Algorithm development requires the long-term assessment of potential biases, which Ahmed et al. (2020) identify as essential for resolving transparency issues before deploying the technology widely. Stakeholders and patients must have confidence and accountability if they are to rely on AI-produced decisions. It is recommended that healthcare leaders establish specific parameters for detecting bias which include factors beyond population demographics such as socioeconomic status, geographic location, and access to care which should also be considered for a comprehensive bias-mitigation process. By recognizing overlooked factors healthcare leaders can ensure algorithms operate equitably without unintentionally marginalizing certain patient groups. Overcoming vital biases proves crucial to establish trust and confidence among patients regarding AI-powered healthcare choices by showing a dedication to technology accountability and fair practice.

Cultural Competency Training

Providing healthcare professionals with the necessary tools to identify and counter bias is also important. As different interviewees argue, “AI bias does not just motivate us to do wrong. It cements itself as another way to systemically do wrong to people society already has it out for.” Predatory inclusion becomes apparent when healthcare teams involve underrepresented groups in AI discussions yet ignore their valuable perspectives. Cultural competency initiatives may end up being merely symbolic unless there is authentic dedication to closing gaps between different groups.

In this case, AI bias demands cultural competency training that is thorough enough for healthcare teams to identify and proactively uncouple all biases. However, according to Barber et al. (n.d), there are implicit biases among healthcare providers, which can unconsciously affect AI outputs if clinicians rely excessively on algorithm recommendations. Case studies and simulations should be used in cultural competency training programs that show how biased AI affects real-life situations. For instance, training modules should show examples of algorithms that disproportionately send resources to majority populations and leave minority groups underserved. This would also encourage a better understanding of how biases exist in healthcare.

Moreover, these programs should be focused on interdisciplinarity. Training initiatives benefit by involving stakeholders from underserved backgrounds to address systemic disparities. When asked, interview participants explained, "Equity starts with understanding your role in the process." This illustrates the need for a holistic approach in which everyone in the healthcare ecosystem participates to minimize disparities. It is recommended that healthcare leaders prioritize cultural competency training programs, which equip professionals with essential tools to identify and mitigate AI system biases. This can be done through healthcare leaders creating training programs that are broad in scope and integrate real-world case studies and simulations to educate staff about bias effects on patient treatment. It is important that leaders who have a voice in the creation and or the utilization of these systems advocate for an interdisciplinary approach that includes stakeholders from underserved communities to integrate diverse perspectives and effectively tackle systemic disparities.

Transparent Governance Frameworks

In healthcare, transparency is necessary to ensure ethical AI implementation. The various interviews highlighted a call for clear policies that bind AI decision-makers to the consequences of their actions. According to the interviewees, because healthcare is risk-averse, ethical AI needs a framework for risk mitigation as an additional protection. Bernal and Mazo (2022) back this by reinforcing transparency's importance in building trust and fairness in AI-driven healthcare systems.

Transparency can be encouraged by conducting regular audits. In addition, public reporting shares information that one can use to audit the algorithmic performance. These measures help organizations find patterns of bias and correct them as soon as possible. Frameworks should also be developed with input from people who have expertise working with underrepresented populations. Patient, clinical, and data science advisory boards may all have interesting perspectives on ethical considerations and equity.

Transparency is also a legislative issue. As evidenced by Singh (2024), the European Union's AI Act is a good model for developing firm rules to check and correct algorithmic bias. Thus, global policies should be adopted, but care should be taken to ensure their compatibility with the healthcare system. Regulators ought to calibrate this oversight to ensure that the users of AI applications are also directly accountable to the populations they are intended to serve. True transparency requires more than token representation. Predatory inclusion in governance happens when decision-makers include marginalized voices in discussions without following through with meaningful policies to address their concerns. Ethical AI governance needs to transcend mere visibility by enabling these communities to actively participate in creating fair AI policies.

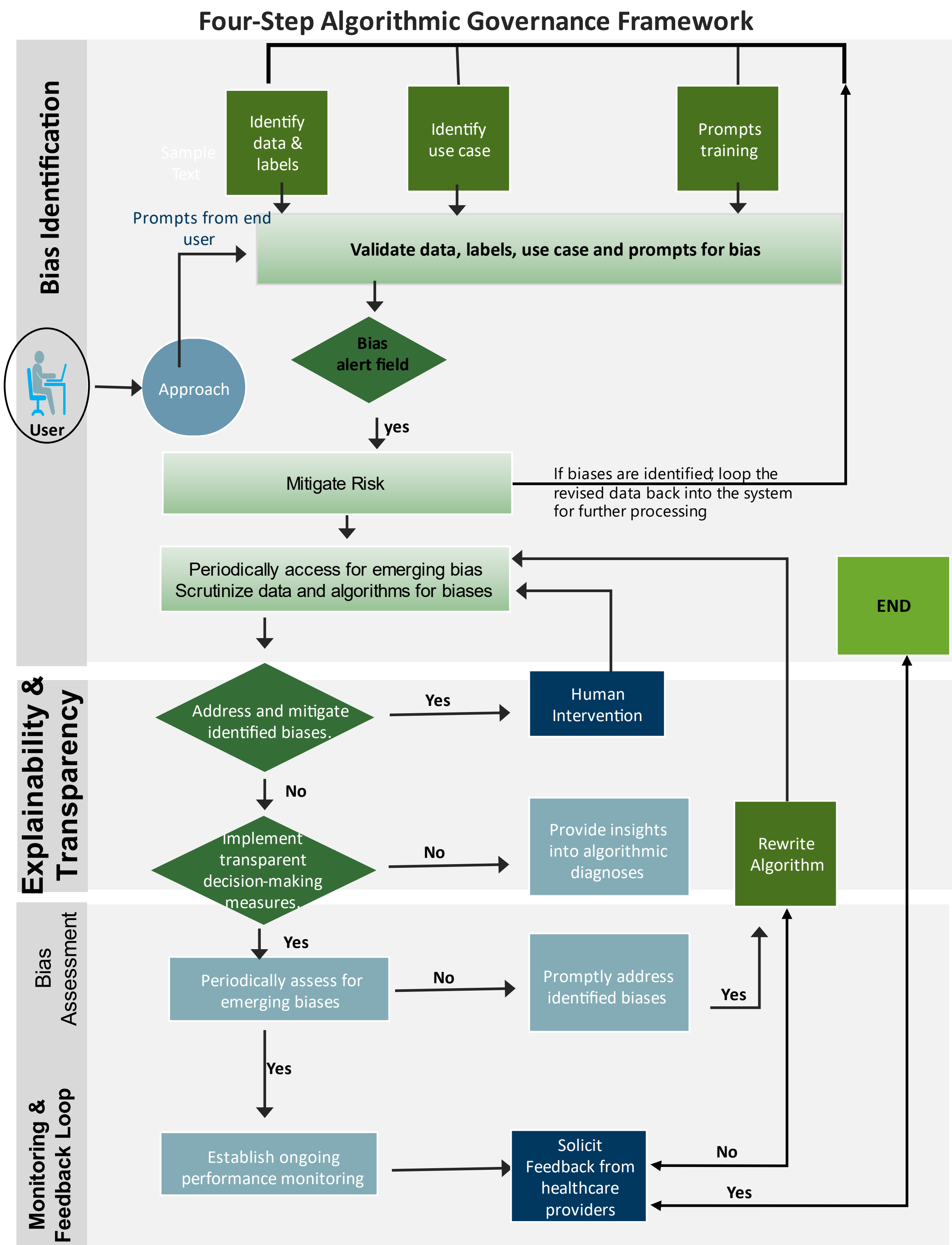
Ethical AI Governance and Leadership Roles

As shown in the literature, the demand for powerful leadership and effectual governance is increasing to reduce AI bias in providing healthcare. Instead of being an issue left to data scientists, leaders are being called forward to advocate for cross-functional collaboration with clinicians, ethicists, scientists, and policymakers to address bias adequately. Leadership should fight for accountability, inclusiveness, and equity within an organization's members. The results will spur said leaders to champion structural changes when creating rules and regulations and strongly encourage the design and implementation of fair AI.

As one of the interviewees noted, "Leadership is just very much about giving a clear signal about what accountability is and how you create an inclusive organizational culture." Furthermore, leaders should also value empowerment efforts to train and equip healthcare teams with tools to understand and manage AI bias. AI technology should also be made to operate within ethical and human-centered values, so that the leaders championing these efforts can build a trusting culture. The governance of ethical AI is as much an organizational responsibility as a technical problem for leaders who decide to develop technical platforms efficiently while creating equity and fairness in healthcare outcomes.

Figure *7*

Four-Step Algorithmic Governance Framework Process Flow



Source: *Shaw (2024). Image illustrating Algorithmic Governance Framework Process Flow*

An example of such a process is depicted in figure seven above. In the context of the healthcare provider example provided, algorithmic governance plays a crucial role in safeguarding against potential biases, ensuring transparency, and promoting accountability throughout the decision-making process (see Appendix A). Healthcare leaders should integrate a robust ethical AI governance framework into their AI roadmap to maintain transparency, accountability, and fairness in AI systems. It is recommended that organizations create specific teams or roles dedicated to AI ethics and governance work so these tasks are not assigned to current leadership positions or an extension of current roles. The effectiveness of AI systems monitoring and management depends on leaders who prioritize ethical standards. Prioritizing data transparency and traceability allows organizations to map AI decisions back to the originating data documents and the understanding of how these decisions were made. This approach reinforces both accountability and adherence to ethical standards. Leaders need to identify whether issues stem from data errors or AI bias and establish specific processes to handle each situation appropriately.

The recommendation is to implement a four-step governance framework as the one above (figure six) to identify and address biases by first examining all data sets including training data alongside algorithms for potential biases and fix or rewrite biased components within the system. I recommend that leaders build cross-disciplinary teams with healthcare providers and patient representatives along with external stakeholders to develop AI systems that represent diverse needs and viewpoints. AI system refinement and ethical standards maintenance require active collaboration. Leaders should explore automated bias mitigation systems based on European Union models while reserving human oversight for final assessments. Healthcare providers and patients should provide ongoing feedback to maintain AI systems' accountability and transparency while adhering to fairness and equity principles. Implementing these recommendations enables healthcare organizations to establish ethical and reliable AI systems that enhance patient care results.

Algorithmic Governance

Algorithmic governance refers to the set of policies, procedures, and mechanisms put in place to ensure that algorithms are developed, deployed, and utilized in a responsible and ethical manner. An example of such a process is depicted in figure six above. In the context of the healthcare provider example provided, algorithmic governance plays a crucial role in safeguarding against potential biases, ensuring transparency, and promoting accountability throughout the decision-making process (see Appendix A). The conceptual framework for this paper incorporates the following elements of algorithmic governance:

Bias Identification

Scrutinize the data and algorithms for potential biases: During this stage biases originating from both data characteristics and algorithmic structure are detected and corrected at an early stage.

Address and mitigate identified biases: The healthcare provider proves its dedication to equitable patient care through bias identification and reduction, which reflects key algorithmic governance principles.

Explainability and Transparency

Implement measures to make the algorithm's decision-making process transparent: Algorithmic governance relies on transparency so that healthcare providers can track the decision-making process of the algorithm and maintain accountability.

Provide insights into how the algorithm arrives at its diagnoses: The healthcare provider builds trust and boosts collaboration between healthcare professionals and the algorithm by making its decision-making process understandable, which are vital components of algorithmic governance.

Bias Assessment

Periodically assess the algorithm for any emerging biases: The practice of constant bias evaluation supports algorithmic governance principles which limit unbiased algorithm performance over time.

Implement mechanisms to promptly address identified biases: Timely resolution of detected biases shows dedication to ethical algorithmic governance while reducing the risk of patient harm.

Monitoring and Feedback Loop

Establish a system for ongoing monitoring of the algorithm's performance and outcomes: The essential component of algorithmic governance lies in the continuous monitoring which maintains the algorithm's alignment with both ethical standards and regulatory requirements. Solicit feedback from healthcare providers: The healthcare providers improves accountability and responsiveness within algorithmic decision-making by obtaining feedback from healthcare technical staff and providers, thus reinforcing algorithmic governance principles.

Healthcare providers who integrate these components into their policy workflow diagrams can maintain algorithmic governance principles while making sure their algorithms operate ethically and responsibly during development and deployment stages. The healthcare provider strengthens patient trust in algorithmic diagnostics by maintaining fairness and transparency while ensuring accountability during the decision-making process which leads to better patient outcomes and supports algorithmic governance principles in healthcare (see Appendix A for a Sample Framework using a Healthcare Provider Example) to deter bias.

Bias detection requires thorough examination of datasets and training methods along with algorithm analysis to identify and resolve biases at the earliest stage. Bias in datasets emerges from incomplete or unrepresentative data while bias in algorithms may originate from model designs, data processing methods, or the training of algorithms fed with biased training examples. Healthcare providers that work to rectify these biases show their dedication to delivering fair and equal treatment to all patients. Algorithmic governance relies on explainability and transparency to enable healthcare providers to trace the impact of data selection, algorithm design choices, and training decisions on algorithmic results while guaranteeing accountability. Understanding the decision-making process of the algorithm helps healthcare professionals build trust with it and work collaboratively to manage biases present in data and training methods. Periodic evaluations of the algorithm should be conducted to detect and amend biases from data sources, algorithm design or training procedures paired with implementing quick response mechanisms to new biases to minimize patient risk. Ongoing monitoring combined with feedback maintains ethical compliance and regulatory adherence while provider feedback enhances accountability and responsiveness. Healthcare providers who integrate these components into their AI governance policy frameworks can limit biases from data, algorithms, and training processes while upholding ethical AI deployment standards which in turn will boost patient trust and lead to better patient outcomes.

Enhancing Data Transparency through Traceability

Traceability enhances transparency and accountability, which are essential for fostering trust in healthcare AI systems. To achieve data transparency in healthcare, patients and providers must know how algorithms use the data they yield. Several interesting proposals may contribute to this trust, including enabling external audits and limiting the openness of data practices. Sharing datasets for peer review to support algorithm accountability and fairness may be served by anonymizing datasets. However, that also means communicating transparently with patients, because they have a right to know how algorithms use their data. Traceability plays a critical role here, as it allows healthcare organizations to trace decisions back to the specific data used, ensuring accountability and fostering trust with patients. For this, healthcare organizations can adopt initiatives such as public dashboards that predict the accuracy and bias detection rates of healthcare organizations. These tools make it possible to keep algorithmic performance accessible and understandable for stakeholders. Reporting mechanisms also encourage transparency, enabling patients to commit to the ethical use of AI.

Organizations should also develop clear policies on how data will be used; everyone in the AI departments should bear responsibility in this. Making data transparent is a legislative matter. Leadership reinforcement of the European Union's AI Act—a strong model for detecting and mitigating algorithmic bias—is recommendable for healthcare systems worldwide. In addition, leaders could ensure that advisory boards consist of a wide range of stakeholders who provide high-quality input in ethics and equity considerations in AI systems (Vashishth et al., 2024). Therefore, making the data available allows healthcare organizations to construct AI-based ethical systems and care delivery that is equitable and patient-centered (Nazer et al., 2023).

The healthcare system's inefficiencies and biases can be fixed only if the education sectors, industry, practitioners, and the US Government recommend inclusive algorithms, cultural competency training, clear governance frameworks, and ethical leadership. Healthcare-related AI is all about bridging humanity with technology. As a result, healthcare leaders can develop equally innovative and faithful AI systems by imitating these strategies. It is recommended that healthcare leaders focus on deploying traceability systems to make AI decision-making processes transparent and accountable. Leaders must also work to establish policies which define data usage steps while maintaining transparency throughout the AI development process, reinforcing their legal frameworks and creating diverse advisory boards to develop equitable AI systems focused on patient needs which connect technology with human healthcare services.

Limitation and Reflection on Research Bias

Through this study, healthcare leaders gain valuable insights into how to tackle algorithmic bias. Nevertheless, the study is not without limitations. A key constraint of much of this work is that the geographic focus is on healthcare leaders from Delaware and the East Coast, which does not fully account for the diversity of challenges healthcare leadership faces in other parts of the country. For example, healthcare systems in rural or underdeveloped areas are usually under-resourced and struggle with certain cultural or demographic problems. A larger scope of geography could permit a wider view into how algorithmic bias emerges and how it is approached and perceived in different healthcare systems.

Also, the study findings may be limited by the sample size. The main advantage of qualitative research is its capacity for depth, but the relatively small number of respondents generally limits the generalizability of findings. Unlike IPA, which is perfect for investigating lived experiences, the IPA method of study profoundly depends on the depth of individual narratives rather than broader quantitative research. The findings are nuanced, yet the insights may differ across the wider population of healthcare leaders. Future investigations might use mixed methods to intermingle the rigor of qualitative data with a survey, which is generalizable quantitative data.

Researcher bias is another concern. Although the researchers strove for objectivity, their perspectives inevitably affected IPA analysis. Considering this, efforts were made to ensure rigor via triangulation and member checking, but possibly some bias infiltrated the findings. Interviews also served as the main data source, and this comes with the potential for response bias; participants may provide socially desirable responses rather than candid reflections. In addition, using literature to frame the study implies temporal limitations. The findings and recommendations are rapidly influenced by the rapid progression of artificial intelligence and healthcare technologies. This demonstrates, however, that the work is just getting started when it comes to keeping up with the shifting landscape of AI in healthcare. Broad, diverse sampling, a combination of quantitative methods, and the imperative to maintain a dynamic research agenda will improve the robustness and relevance of future studies addressing these limitations.

Future Research Directions

After finishing my interviews, I realized that future research could benefit from quantitative analysis about how healthcare leaders affect AI bias reduction in decision-making processes. This research would study how leadership initiatives such as AI ethics teams' formation, training program development, and routine AI audits help identify and reduce bias in healthcare outcomes produced by AI systems. The study will collect data from healthcare leaders, professionals, and patients through surveys and questionnaires while analyzing AI decision outputs to identify demographic disparities and other factors. The evaluation of leadership's impact on AI bias resolution can be achieved through the application of statistical methods such as regression analysis and comparative studies. Healthcare leaders can apply optimal AI management practices revealed by this research to develop ethical and fair AI-use policies and operational strategies.

The results from this study also identified important avenues for tackling algorithmic bias in underserved regions of the world and cultivating better patient inclusion in the use of AI in healthcare. In the future, further examinations of how AI impacts healthcare delivery in low-resource settings will reveal more. Some regions have poor health disparities and limited access to advanced technology. Understanding how AI could bridge or improve equitable healthcare innovation in that context is important. Another area where investigation is needed is the patient perspective of AI in the clinical setting. Though this study focuses on healthcare leaders' voices, it overlooks patient perspectives, especially those of marginalized community members. What they experience with AI-driven decisions should be prioritized in future research, especially regarding how much they trust, understand, and perceive those technologies. Since this is not just a technical issue but a human one, patients need to see and understand the logic these tools are based on.

Moreover, longitudinal studies are needed to evaluate the long-term effects of using AI bias mitigation strategies. Such studies would allow us to determine the required sustainable interventions, how biases evolve, and the broader impact of AI on healthcare outcomes. Similarly, bias detection libraries could be used through pilot programs to track success in reducing disparities in care provision. Additionally, future research should focus on interdisciplinary collaboration. Having ethicists, data scientists, and sociologists engage with healthcare professionals will provide a holistic view of how bias works and how to counter it. According to the literature, "Bias is not just a technical problem; it is a cultural and systemic one" (Bernal & Mazo, 2022). Therefore, research efforts should go beyond technical solutions to include cultural competency training, policy development, and stakeholder engagement.

Also, given its unparalleled capability to process massive datasets in tandem with one another, quantum computing is a transformative resource in healthcare when it is melded with AI and machine learning. Quantum algorithms enable AI systems to find patterns and insights that are otherwise computationally lethal, eventually leading to breakthrough diagnostics, treatment optimization, and personalized medicine. Quantum-enhanced drug discovery and diagnosis through AI may improve these outcomes substantially. However, as Singh (2024) suggests, it could also be the means for greater inequity and reduced accessibility. Thus, in the future, care should be taken while navigating these technologies so that their use can address disparities rather than exacerbate them. Integrating quantum AI in healthcare decision-making requires strong ethical guidelines and leadership to improve patients' equity and inclusivity. The option of quantum computing and AI as a future-forward synergy highlights the importance of ongoing interdisciplinary work to equitably reduce algorithmic bias and promote alternative healthcare solutions.

Furthermore, healthcare facilities must explore incorporating emotional intelligence (EQ) into AI systems to develop novel ways to reduce bias in AI systems. According to Monlezun, AI EQ is worthy of further investigation to determine how it could improve AI's capability to make empathetic decisions in applications involving patients (2023). The choice of research directions will impact what we can explore and achieve with AI in the next few years. Focusing on these avenues will help move the field toward more equitable, ethical, and successful deployment of AI in healthcare.

Conclusion

The findings reveal significant ramifications of algorithmic bias in healthcare, from clinical outcomes to organizational workflows and trust in technology. The urgency of action cannot be overstated at this stage of AI integration into healthcare systems. As an antidote to bias, healthcare leaders play an essential role. This research calls attention to the importance of transparent, inclusive, and accountable practices in achieving such goals.

First, algorithmic bias is a product of historical and systemic inequities that are deeply entrenched: "the biases of its time." Thus, not only do we need technical interventions to address these biases, but we also need to effect cultural change within organizations. This includes creating a common knowledge of equity, focusing on multiple and diverse data collection, and urging cooperation across disciplines. The research also has important implications for ethical development. AI in the healthcare context requires holistic governance frameworks with characteristics of fairness and accountability from all the stages of the life cycle.

Bias within healthcare systems extends beyond its historical data origins and represents a foundational problem that affects patient results and fairness, not to mention the reliability of AI decision-making tools. This paper demonstrates that bias represents both an obstacle and a potential advantage. If bias remains unmonitored, it will continue to cause differences in healthcare outcomes, but when it is addressed with purpose, it transforms into a means to eliminate healthcare disparities. AI functions as a double-edged sword because it both mirrors and intensifies existing biases found in its training data.

With these findings in mind, bias detection tools should be put in place, data transparency should be ensured, and underserved communities should be considered as part of an AI system from the outset to improve their confidence and trust in the system *before* its implementation. Moreover, AI applications should not be used in any context unless healthcare professionals are trained in cultural competency to ensure the applications used appropriately for varied patient needs. Healthcare organizations need to implement forward-thinking bias mitigation strategies that incorporate education, ethical AI governance principles, and inclusive algorithm design to address AI-driven decision-making challenges. Healthcare organizations need to maintain transparency and accountability as their highest priorities to guarantee AI systems deliver efficiency alongside fairness and patient-centered care.

Successful implementation of AI in healthcare requires changing how we view bias—from something unavoidable to something that we must responsibly control. Organizations should promote teamwork across functions while embedding cultural understanding and regularly updating their frameworks for AI supervision. Healthcare innovation requires sustained dedication to ethical practices that promote equity and intelligence. To foster sustainable innovation and operational excellence healthcare leaders must employ both top-down and bottom-up approaches in their AI governance strategies. Executive leadership should create a strategic AI governance framework from the top down that focuses on transparency, data control, and interoperability to achieve effective intellectual property management while maintaining output control and vendor independence. Organizations without this foundation face regulatory challenges and operational inefficiencies which can lead to a loss of competitive advantage.

Healthcare leaders need to enable their teams to execute AI projects with defined supervision and ethical protection while developing flexible systems that connect smoothly across various platforms. The combination of model transparency together with flexible architecture and data ownership allows organizations to manage AI challenges while meeting compliance standards and achieving long-term success in the dynamic healthcare sector.

Further research is necessary to adapt these strategies iteratively and look for alternative innovation paths. By laying the foundation for further studies with a more expansive geographic and demographic focus, the inclusion of patient experiences, and interdisciplinary collaboration, this work can serve as a model for health disparities in other parts of the country. The healthcare landscape continues to change, influencing how we ensure ethical, equitable, and effective AI implementation. Overall, this study contributes to the burgeoning conversation around AI ethics and equity in healthcare by studying the experiences of healthcare leaders to extract critical insights into the challenges and opportunities of addressing algorithmic bias.

As AI technology continues to shape the future of medicine, the call to action is clear: Bias must not be an afterthought. System design must begin with eliminating bias to create equitable healthcare solutions for every patient from any racial, gender, or socioeconomic group. To unleash AI's transformative potential, all stakeholders, leaders, policymakers, technologists, and communities must come together to build a path toward a better future. The true value of healthcare AI will come from generating significant and inclusive improvements in patient health outcomes, rather than just achieving high accuracy levels. At the same time, vigilance and cooperation in the ongoing process can help ensure that technological advancements regarding AI lead to better outcomes for all.

References

Agency for Healthcare Research and Quality. (2023). *Primary care and patient safety: Opportunities at the interface*. https://psnet.ahrq.gov/perspective/primary-care-and-patient-safety-opportunities-interface

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

AI Factory. (2024, December 20). *The impact of bias in AI algorithms on patient outcomes*. AI Factory. https://aifactory.technology/the-impact-of-bias-in-ai-algorithms-on-patient-outcomes

Aisha Shalash, Niveen M E Abu-Rmeileh, Dervla Kelly, & Khalifa Elmusharaf. (2022b). The need for standardised methods of data collection, sharing of data and agency coordination in humanitarian settings. *BMJ Global Health*, *7*(Suppl 8), e007249. https://doi.org/10.1136/bmjgh-2021-007249

Aquino, Y. S. J., Carter, S. M., Houssami, N., Braunack-Mayer, A., Win, K. T., Degeling, C., ... & Rogers, W. A. (2023). Practical, epistemic and normative implications of algorithmic bias in healthcare artificial intelligence: a qualitative study of multidisciplinary expert perspectives. Journal of Medical Ethics.

Arora, A., Alderman, J. E., Palmer, J., Shaswath Ganapathi, Laws, E., McCradden, M. D., Oakden-Rayner, L., Pfohl, S. R., Ghassemi, M., McKay, F., Treanor, D., Negar Rostamzadeh, Mateen, B., Gath, J., Adebajo, A. O., Kuku, S., Matin, R., Heller, K., Sapey, E., & Sebire, N. J. (2023). The value of standards for health datasets in artificial intelligence-based applications. *Nature Medicine*, *29*(11), 2929–2938. https://doi.org/10.1038/s41591-023-02608-w

Avanceña, A. L., & Prosser, L. A. (2021). Examining equity effects of health interventions in cost-effectiveness analysis: a systematic review. Value in Health, 24(1), 136-143.

Barber, D. H., Wilson, T., Vrablik, L., & Wing, R. (n.d.). Implicit Bias and Patient Care: Mitigating Bias, Preventing Harm. *MedEdPORTAL*, *19*, 11343. https://doi.org/10.15766/mep_2374-8265.11343

Bernal, J., & Mazo, C. (2022). Transparency of Artificial Intelligence in Healthcare: Insights from Professionals in Computing and Healthcare Worldwide. *Applied Sciences*, *12*(20), 10228–10228. https://doi.org/10.3390/app122010228

Bhargava, C., & Sharma, P. K. (2021). *Artificial Intelligence: Fundamentals and Applications*. Taylor & Francis Group. http://ebookcentral.proquest.com/lib/brandeis-ebooks/detail.action?docID=6644860

Bias in Algorithms – Artificial intelligence and discrimination. (n.d.).

Bini, S. A. (2018). Artificial Intelligence, Machine Learning, Deep Learning, and Cognitive Computing: What Do These Terms Mean and How Will They Impact Health Care? The Journal of Arthroplasty, 33(8), 2358-2361. https://doi.org/10.1016/j.arth.2018.02.067

Barton, M., Hamza, M., & Guevel, B. (2023). Racial equity in healthcare machine learning: Illustrating bias in models with minimal bias mitigation. Cureus, 15(2), e35037. https://doi.org/10.7759/cureus.35037

Belenguer, L. (2022). AI bias: Exploring discriminatory algorithmic decision-making models and the application of possible machine-centric solutions adapted from the pharmaceutical industry. *AI and Ethics, 2*(4), 771–787. https://doi.org/10.1007/s43681-022-00138-8

Celi, L. A., Cellini, J., Charpignon, M. L., Dee, E. C., Dernoncourt, F., Eber, R., ... & Yao, S. (2022). Sources of bias in artificial intelligence that perpetuate healthcare disparities—A global review. PLOS Digital Health, 1(3), e0000022.

Chapman, E., & Smith, J. A. (2002). Interpretative phenomenological analysis and the new genetics. Journal of health psychology, 7(2), 125-130.

Char, D. S., Abràmoff, M. D., & Feudtner, C. (2020). Identifying ethical considerations for machine learning healthcare applications. The American Journal of Bioethics, 20(11), 7-17.

MCheng, J. Y., Abel, J. T., Balis, U. G., McClintock, D. S., & Pantanowitz, L. (2021). Challenges in the development, deployment, and regulation of artificial intelligence in anatomic pathology. The American Journal of Pathology, 191(10), 1684-1692.

Cho, M. K. (2021a). Rising to the challenge of bias in health care AI. *Nature Medicine*, *27*(12), 2079–2081. https://doi.org/10.1038/s41591-021-01577-2

Cohn, L. (2024, December 13). *AI regulation in 2025: A comparative guide to U.S. state and EU rules*. Gartner. https://www.gartner.com/document-reader/document/6015603?ref=TypeAheadSearch

Collins, C. (2024, May 7). AI and algorithms: The other side of medical claims processing. American Institute of Healthcare Compliance. https://aihc-assn.org/ai-and-algorithms-the-other-side-of-medical-claims-processing/

ComplexDiscovery Staff. (2024, December 7). AI in healthcare: UnitedHealthcare's controversy and the tragic death of CEO. ComplexDiscovery. https://complexdiscovery.com/ai-in-healthcare-unitedhealthcares-controversy-and-the-tragic-death-of-ceo/

Crigger, E., & Khoury, C. (2019). *MEDICINE AND SOCIETY*.

Cross, J. L., Choma, M. A., & Onofrey, J. A. (2024). Bias in medical AI: Implications for clinical decision-making. PLOS Digital Health, 3(11), e0000651. https://doi.org/10.1371/journal.pdig.0000651

Data and Organisations - Leuphana University (Director). (2021, October 4). *Algorithmic Governance* [Video recording]. https://www.youtube.com/watch?v=vVitCrDu8ms

Defining the PCMH. (n.d.). [Federal Healthcare Agency]. The Agency for Healthcare Research and Quality's (AHRQ). Retrieved June 3, 2024, from https://www.ahrq.gov/ncepcr/research/care-coordination/pcmh/define.html

Difference between AI vs Machine Learning vs Deep Learning. (n.d). Jelvix. Retrieved April 23, 2024, from https://jelvix.com/blog/ai-vs-machine-learning-vs-deep-learning

Djap, S. (2022, February 27). *Defining algorithmic bias*. Center for Critical Race and Digital Studies. https://www.criticalracedigitalstudies.com/peoples-guide-posts/defining-algorithmic-bias

Dlugatch, R., Georgieva, A., & Kerasidou, A. (2024). AI-driven decision support systems and epistemic reliance: A qualitative study on obstetricians' and midwives' perspectives on integrating AI-driven CTG into clinical decision making. *BMC Medical Ethics*, *25*, 1–11. https://doi.org/10.1186/s12910-023-00990-1

Excerpt from Rainbow in the Cloud. (n.d.). Penguin Random House Canada. Retrieved April 5, 2024, from https://www.penguinrandomhouse.ca/books/239371/rainbow-in-the-cloud-by-maya-angelou/9780812996456/excerpt

Explainable AI. (n.d.). Data Robot AI Platform. Retrieved May 1, 2024, from https://www.datarobot.com/wiki/explainable-ai/

Forum, W. E. (n.d.). *Strategic Intelligence | World Economic Forum*. Stategic Intelligence.

Retrieved May 15, 2024, from https://intelligence.weforum.org

Gartner. (2024). What Generative AI Means for Business. https://gcom.pdo.aws.gartner.com/en/insights/generative-ai-for-business

Glossary Resources: Designing and Building AI Products and Services. (n.d.). Retrieved July 1, 2024, from https://classroom.emeritus.org/courses/5972/pages/glossary-resources?module_item_id=1671541

Hacker, P., Engel, A., & Mauer, M. (2023). Regulating ChatGPT and other Large Generative AI Models. *Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency*, 1112–1123. https://doi.org/10.1145/3593013.3594067

Haider, S. A., Borna, S., Gomez-Cabello, C. A., Pressman, S. M., Haider, C. R., & Forte, A. J. (2024). The Algorithmic Divide: A Systematic Review on AI-Driven Racial Disparities in Healthcare. Journal of Racial and Ethnic Health Disparities. https://doi.org/10.1007/s40615-024-02237-0

Hall, W. J., Chapman, M. V., Lee, K. M., Merino, Y. M., Thomas, T. W., Payne, B. K., Eng, E., Day, S. H., & Coyne-Beasley, T. (2015). Implicit Racial/Ethnic Bias Among Health Care Professionals and Its Influence on Health Care Outcomes: A Systematic Review. *American Journal of Public Health*, *105*(12), e60-76. https://doi.org/10.2105/AJPH.2015.302903

Hanna, M., Liron Pantanowitz, Jackson, B., Palmer, O., Shyam Visweswaran, Pantanowitz, J., Mustafa Deebajah, & Rashidi, H. (2024). Ethical and Bias Considerations in Artificial Intelligence (AI)/Machine Learning. *Modern Pathology*, 100686–100686. https://doi.org/10.1016/j.modpat.2024.100686

Health Data, Technology, and Interoperability: Certification Program Updates, Algorithm

Transparency, and Information Sharing (HTI-1) Final Rule | HealthIT.gov. (n.d.). Retrieved May 11, 2024, from https://www.healthit.gov/topic/laws-regulation-and-policy/health-data-technology-and-interoperability-certification-program

HealthITSecurity. (2023, December 20). *How Will Biden's Executive Order on Trustworthy AI Impact Healthcare?* HealthITAnalytics. https://healthitanalytics.com/features/how-will-bidens-executive-order-on-trustworthy-ai-impact-healthcare

Helm, J. M., Swiergosz, A. M., Haeberle, H. S., Karnuta, J. M., Schaffer, J. L., Krebs, V. E., ... & Ramkumar, P. N. (2020). Machine learning and artificial intelligence: definitions, applications, and future directions. Current reviews in musculoskeletal medicine, 13, 69-76.

Hennink, M. M., Kaiser, B. N., & Marconi, V. C. (2017). Code Saturation Versus Meaning Saturation: How Many Interviews Are Enough? *Qualitative Health Research*, *27*(4), 591–608. https://doi.org/10.1177/1049732316665344

Https://search.library.brandeis.edu. (n.d.). Retrieved April 2, 2024, from https://search.library.brandeis.edu

Hoffman, S., & Podgurski, A. (2020). *Artificial intelligence and discrimination in health care*. Yale Journal of Health Policy, Law, and Ethics, 19(1), Article 3. https://digitalcommons.law.yale.edu/yjhple/vol19/iss1/3

IBM. (n.d.). *What is machine learning (ML)?* IBM. https://www.ibm.com/topics/machine-learning

Jones, S., Scott, E., & Williams, T. (2025). Groundbreaking AI executive order issued. *Jones Day*. https://www.jonesday.com/en/insights/2023/10/groundbreaking-ai-executive-order-issued

Joubert, S. (2023, August 8). *Effective Leadership in Healthcare: 5 Essential Traits*. Graduate Blog. https://graduate.northeastern.edu/resources/leadership-in-healthcare/

Kanade, V. (2022, April 4). *What is machine learning? Definition, types, applications, and trends*. Spiceworks. https://www.spiceworks.com/tech/artificial-intelligence/articles/what-is-ml/

Kasthuri, N., & Meeradevi, T. (2022). AI-Driven Healthcare Analysis. In *Smart Systems for Industrial Applications* (pp. 269–285). John Wiley & Sons, Ltd. https://doi.org/10.1002/9781119762010.ch11

Kim, J. Y., Hasan, A., Kellogg, K., Ratliff, W., Murray, S., Suresh, H., Valladares, A., Shaw, K., Tobey, D., Vidal, D. E., Lifson, M. A., Patel, M., Raji, I. D., Gao, M., Knechtle, W., Tang, L., Balu, S., & Sendak, M. P. (2023). Development and preliminary testing of Health Equity Across the AI Lifecycle (HEAAL): A framework for healthcare delivery organizations to mitigate the risk of AI solutions worsening health inequities. *medRxiv*. https://doi.org/10.1101/2023.10.16.23297076

Krautscheid, L. C. (2014). Defining Professional Nursing Accountability: A Literature Review. *Journal of Professional Nursing*, *30*(1), 43–47. https://doi.org/10.1016/j.profnurs.2013.06.008

Larson, D. B., Harvey, H., Rubin, D. L., Irani, N., Justin, R. T., & Langlotz, C. P. (2021). Regulatory frameworks for development and evaluation of artificial intelligence–based diagnostic imaging algorithms: summary and recommendations. Journal of the American College of Radiology, 18(3), 413-424.

Laukka, E., Huhtakangas, M., Heponiemi, T., & Kanste, O. (2020). Identifying the Roles of Healthcare Leaders in HIT Implementation: A Scoping Review of the Quantitative and

Qualitative Evidence. *International Journal of Environmental Research and Public Health*, *17*(8), 2865. https://doi.org/10.3390/ijerph17082865

Lavin, A., Gilligan-Lee, C. M., Visnjic, A., Ganju, S., Newman, D., Ganguly, S., Lange, D., Baydin, A. G., Sharma, A., Gibson, A., Zheng, S., Xing, E. P., Mattmann, C., Parr, J., & Gal, Y. (2022). Technology readiness levels for machine learning systems. *Nature Communications*, *13*(1). https://doi.org/10.1038/s41467-022-33128-9

Levin, B., & Downes, L. (2023, May 19). Who Is Going to Regulate AI? *Harvard Business Review*. https://hbr.org/2023/05/who-is-going-to-regulate-ai

Li, J., Dada, A., Kleesiek, J., & Egger, J. (2023). *ChatGPT in Healthcare: A Taxonomy and Systematic Review* (p. 2023.03.30.23287899). medRxiv. https://doi.org/10.1101/2023.03.30.23287899

Locke, T., Parker, V. J., Thoumi, A., Goldstein, B. A., & Silcox, C. (2023). *Preventing Bias and Inequities in AI-Enabled Health Tools*. Duke-Margolis Center for Health Policy.

Marques, T. J. F. (2021). Overcoming algorithmic bias: the role of bias awareness, knowledge, and minority status on human decision-making (Doctoral dissertation).

Medical Algorithms. (n.d.). Medical chart review checklist. Medical Algorithms. Retrieved February 24, 2025, from https://www.medicalalgorithms.com/medical-chart-review-checklist

Mennella, C., Maniscalco, U., Pietro, G. D., & Esposito, M. (2024). Ethical and regulatory challenges of AI technologies in healthcare: A narrative review. *Heliyon*, *10*(4), e26297–e26297. https://doi.org/10.1016/j.heliyon.2024.e26297

Meskó, B., & Topol, E. J. (2023). The imperative for regulatory oversight of large language models (or generative AI) in healthcare. *NPJ Digital Medicine*, *6*(1), 120.

https://doi.org/10.1038/s41746-023-00873-0

Mitchell, T. M. (1997). *Machine Learning*. McGraw Hill.

Monlezun, D. J. (2023). *The Thinking Healthcare System: Artificial Intelligence and Human Equity*. Elsevier.

Nabi, J. (2018). How bioethics can shape artificial intelligence and machine learning. *Hastings Center Report, 48*(5), 10–13. https://doi.org/10.1002/hast.895

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

Network of the National Library of Medicine. (n.d.). Algorithm. NNLM. Retrieved from https://www.nnlm.gov/guides/data-glossary/algorithm

News, H. I. (2023, August 10). *GenHealth.AI Accelerates into Healthcare AI Market with New Funding, Advisor Appointments and Use Cases Across Healthcare*. HospitalEMRandEHR.Com [BLOG]; Newstex. https://www.proquest.com/docview/2848189486/citation/147469D27E394C3FPQ/1

Nikolinakos, N. Th. (2023). Ethical Principles for Trustworthy AI. In N. Th. Nikolinakos (Ed.), *EU Policy and Legal Framework for Artificial Intelligence, Robotics and Related Technologies—The AI Act* (pp. 101–166). Springer International Publishing. https://doi.org/10.1007/978-3-031-27953-9_3

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://www.cell.com/patterns/fulltext/S2666-3899(21)00202-6

Obermeyer, Z., Powers, B., Vogeli, C., & Sendhil Mullainathan. (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

Ocloo, J., & Matthews, R. (2016). From tokenism to empowerment: Progressing patient and public involvement in healthcare improvement. *BMJ Quality & Safety*, *25*(8), 626–632. https://doi.org/10.1136/bmjqs-2015-004839

Olsen, E. (2024). Healthcare technology trends to watch in 2024. *HR Dive*. https://www.proquest.com/docview/2914964425/abstract/5F2CFCA1775144D0PQ/1

Open-source data science: How to reduce bias in AI. (2022, October 14). World Economic Forum. https://www.weforum.org/agenda/2022/10/open-source-data-science-bias-more-ethical-ai-technology/

Panch, T., Mattie, H., & Atun, R. (2019). Artificial intelligence and algorithmic bias: Implications for health systems. *Journal of Global Health, 9*(2), 020318. https://doi.org/10.7189/jogh.09.020318

Pathak, V., Jena, B., & Kalra, S. (2013). Qualitative research. Perspectives in clinical research, 4(3), 192.

Paulus, J. K., & Kent, D. M. (2020). Predictably unequal: understanding and addressing concerns that algorithmic clinical prediction may increase health disparities. npj Digital Medicine, 3(99). https://doi.org/10.1038/s41746-020-0304-9

Phillips, P. J., & Pohl, G. (2020). Countering Intelligence Algorithms: Decision Theory, Design Choices and Counter-AI. *The RUSI Journal*, *165*(7), 22–32. https://doi.org/10.1080/03071847.2021.1893126

Price, T., Tredinnick-Rowe, J., Walshe, K., Tazzyman, A., Ferguson, J., Boyd, A., Archer, J., &

Bryce, M. (2020). Reviving clinical governance? A qualitative study of the impact of professional regulatory reform on clinical governance in healthcare organisations in England. *Health Policy*, *124*(4), 446–453. https://doi.org/10.1016/j.healthpol.2020.01.004

Racial Bias Might Be Infecting Patient Portals. Can AI Help? (2024, May 30). HBS Working Knowledge. http://hbswk.hbs.edu/item/racial-bias-might-be-infecting-patient-portals-can-ai-help

Rayhan, A., Rayhan, R., & Rayhan, S. (2023). *The Role Of AI In Healthcare: Revolutionizing Patient Care And Well-Being*. https://doi.org/10.13140/RG.2.2.22601.93283

Rines, J., Daley, K., Loo, S., Safari, K., Walsh, D., Gill, M., ... & Marshall, D. (2022). A patient-led, peer-to-peer qualitative study on the psychosocial relationship between young adults with inflammatory bowel disease and food. Health Expectations, 25(4), 1486-1497.

Robb, L. (n.d.). *DECISION MAKING POLICY*.

Rojas, J. C., Fahrenbach, J., Makhni, S., Cook, S. C., Williams, J. S., Umscheid, C. A., & Chin, M. H. (2022). Framework for Integrating Equity into Machine Learning Models. *Chest*, *161*(6), 1621–1627. https://doi.org/10.1016/j.chest.2022.02.001

Role of Trust in AI-Driven Healthcare Systems: Discussion from the Perspective of Patient Safety. (n.d.). https://doi.org/10.1177/2327857922111026

Sadiku, M. N., Musa, S. M., Sadiku, M. N., & Musa, S. M. (2021). Augmented intelligence. A Primer on Multiple Intelligences, 191-199.

Samuel, A. L. (1959). Some Studies in Machine Learning Using the Game of Checkers. *IBM Journal of Research and Development, 3*(3), 210–229. https://doi.org/10.1147/rd.33.0210

Sen, J. (Ed.). (2021). *Machine Learning—Algorithms, Models and Applications* (Vol. 7). IntechOpen. https://doi.org/10.5772/intechopen.94615

Seyyed-Kalantari, L., Zhang, H., McDermott, M. B. A., Chen, I. Y., & Ghassemi, M. (2021). Underdiagnosis bias of artificial intelligence algorithms applied to chest radiographs in under-served patient populations. *Nature Medicine*, *27*(12), 2176–2182. https://doi.org/10.1038/s41591-021-01595-0

Shalash, A., Abu-Rmeileh, N. M. E., Kelly, D., & Elmusharaf, K. (2022). The need for standardised methods of data collection, sharing of data and agency coordination in humanitarian settings. *BMJ Global Health*, *7*(Suppl 8), e007249. https://doi.org/10.1136/bmjgh-2021-007249

Shinkunas, L. A., Klipowicz, C. J., & Carlisle, E. M. (2020). Shared decision making in surgery: a scoping review of patient and surgeon preferences. BMC medical informatics and decision making, 20(1), 1-14.

Singh, B. K., & Sinha, G. R. (2022). *Machine Learning in Healthcare: Fundamentals and Recent Applications*. Taylor & Francis Group. http://ebookcentral.proquest.com/lib/brandeis-ebooks/detail.action?docID=6876585

Singh, G. V. (2024). *Quantum machine learning in healthcare: diagnostics and drug discovery*.

Smith, J. A., & Nizza, I. E. (2022). What is interpretative phenomenological analysis? In J. A. Smith & I. E. Nizza, *Essentials of interpretative phenomenological analysis.* (pp. 3–10). American Psychological Association. https://doi.org/10.1037/0000259-001

Sriharan, A., Sekercioglu, N., Mitchell, C., Senkaiahliyan, S., Hertelendy, A., Porter, T., & Banaszak-Holl, J. (2024). Leadership for AI Transformation in Health Care Organization: Scoping Review. *Journal of Medical Internet Research*, *26*, e54556–e54556. https://doi.org/10.2196/54556

Tang, M., Mishuris, R. G., Payvandi, L., & Stern, A. D. (2024). Differences in Care Team

Response to Patient Portal Messages by Patient Race and Ethnicity. *JAMA Network Open*, *7*(3), e242618. https://doi.org/10.1001/jamanetworkopen.2024.2618

The Beryl Institute. (2023). *New to patient experience?* https://theberylinstitute.org/new-to-patient-experience/

Trigo, A., Stein, N., & Belfo, F. P. (2024). Strategies to improve fairness in artificial intelligence: A systematic literature review. *Education for Information*, *40*(3), 323–346. https://doi.org/10.3233/efi-240045

Turing, A. M. (1950). Computing machinery and intelligence. Mind, 59(236), 433–460. https://doi.org/10.1093/mind/LIX.236.433

U.S. Department of Health and Human Services. (2021). *Federal Policy for the Protection of Human Subjects ('Common Rule')*. https://www.hhs.gov/ohrp/regulations-and-policy/regulations/common-rule/index.html

U.S. Food and Drug Administration. (2021). *FDA Releases Artificial Intelligence/Machine Learning Action Plan*. U.S. Food and Drug Administration. https://www.fda.gov/news-events/press-announcements/fda-releases-artificial-intelligencemachine-learning-action-plan

Ueda, D., Taichi Kakinuma, Fujita, S., Koji Kamagata, Yasutaka Fushimi, Ito, R., Matsui, Y., Nozaki, T., Takeshi Nakaura, Noriyuki Fujima, Fuminari Tatsugami, Yanagawa, M., Hirata, K., Yamada, A., Takahiro Tsuboyama, Kawamura, M., Fujioka, T., & Shinji Naganawa. (2024). Fairness of artificial intelligence in healthcare: review and recommendations. *Japanese Journal of Radiology*, *42*(1), 3–15. https://doi.org/10.1007/s11604-023-01474-3

Valentin, G., Nielsen, C. V., Nielsen, A.-S. M., Tonnesen, M., Bliksted, K. L., Jensen, K. T., Ingerslev, K., Maribo, T., & Oestergaard, L. G. (2023). Bridging Inequity Gaps in Healthcare Systems While Educating Future Healthcare Professionals—The Social Health Bridge-Building Programme. *International Journal of Environmental Research and Public Health*, *20*(19), 6837. https://doi.org/10.3390/ijerph20196837

Vayena, E., Blasimme, A., & Cohen, I. G. (2018). Machine learning in medicine: Addressing ethical challenges. *PLoS Medicine*, *15*(11), e1002689. https://doi.org/10.1371/journal.pmed.1002689

Veena, G., Balamurugan, B., Nallakaruppan, M. K., Vijay, A., & Mariofanna, M. (2024). *Approaches to Human-Centered AI in Healthcare*. IGI Global.

Völzer, H., Varacca, D., & Kindler, E. (2005). Defining Fairness. In M. Abadi & L. de Alfaro (Eds.), *CONCUR 2005 – Concurrency Theory* (pp. 458–472). Springer. https://doi.org/10.1007/11539452_35

Vashishth, T., Sharma, V., & Sharma, K. (2024). *Merging Artificial Intelligence (AI) and Machine Learning (ML) With Quantum Technology.*

Wang, C., Liu, S., Yang, H., Guo, J., Wu, Y., & Liu, J. (2023). Ethical considerations of using ChatGPT in health care. Journal of Medical Internet Research, 25, e48009.

What is Bias Mitigation. (n.d.). DataHeroes. Retrieved June 10, 2024, from https://dataheroes.ai/glossary/bias-mitigation/

WHO. (2021). Ethics And Governance of Artificial Intelligence for Health Ethics and Governance of Artificial Intelligence for Health. In *WHO*. https://iris.who.int/bitstream/handle/10665/341996/9789240029200-eng.pdf

Wright, S. (2020). Lived experience. In R. Kitchin, N. Thrift, N. Castree, M. Crang, M. Domosh, K. Anderson, P. Cloke, J. Crampton, B. Graham, C. Hadjimichalis, P. Hubbard, R.

Kearns, M.-P. Kwan, L. Lees, S. McLafferty, A. Paasi, C. Philo, J. Sidaway, K. Willis, & H. Yeung (Eds.), *International Encyclopedia of Human Geography* (2nd ed., pp. 1–6). Elsevier. https://doi.org/10.1016/B978-0-08-102295-5.00074-0

Zhang, J., Budhdeo, S., William, W., Cerrato, P., Shuaib, H., Sood, H., Ashrafian, H., Halamka, J., & Teo, J. T. (2022). Moving towards vertically integrated artificial intelligence development. *Npj Digital Medicine*, *5*(1), 143. https://doi.org/10.1038/s41746-022-00690-x

Zotero (n.d.). Retrieved April 22, 2024, from https://www.zotero.org/

Appendix A: Brief Historical Timeline of AI in Healthcare

Time Period	*Event*
Mid-1950s	A regular objective of early endeavors at clinical master frameworks is to supplant the doctor with a Greek prophet model of clinical dynamic. To make a specialist in a container," fit for questioning the doctor or clinical professional regarding the manifestation of a patient, and to produce a conclusion is the purpose of this line of belief
The late 1960s	Early master frameworks started impressive energy in the field of medication, and the late 1960s brought about an elevated degree of desire.
1970	In medication, a persuasive paper about AI was published by Dr. William B. Schwartz in the New England Journal of Medicine. Accordingly, numerous researchers are pulled in to examine the utilization of software engineering in medication
1970	A joint establishment was set by Harvard Medical School and MIT-related training, research, and administration to cultivate the improvement of well-being. Health, Sciences and Technology (HST) was named the new division. Among numerous projects, HST has offered the preparation in clinical informatics, which is a field firmly identified with AI in medication.
1975	Improvement of MYCIN master framework, a standard-based program for the conclusion of bacterial diseases in the blood (Stanford)
1975	Advancement of the INTERNIST master framework, an analytic guide that consolidates an enormous database of illness/signs relationships with methods for issue definition (Pittsburgh)
1976	Advancement of CASNET (Causal-Associated Network) master framework, which utilizes physiological models for the finding and treatment of eye malady (Rutgers et al.)
1977	Advancement of PUFF master framework for programmed translation of pneumonic capacity tests (Pacific Presbyterian Medical Center, San Francisco)

	In its business structure, the PUFF framework has been sold to several international destinations that are still being used today.
The late 1970s	Advancement of PRESENT-ILLNESS master framework, to analyze kidney ailments (MIT). It utilizes a PC model, which is more complex than the standard-based frameworks used in different ventures, an "outline-based" model where an edge-like structure inside the program speaks to living things
1979	Foundation of the American Association for Artificial Intelligence (AAAI)
1981	Improvement of the ABEL master framework, a program that utilizes staggered pathophysiologic models for the conclusion of corrosive base and electrolyte issues (MIT)
Early-1980s	Simulated intelligence in medication is to a great extent US-based research network. Work starts out of numerous grounds, including MIT, Pittsburgh, Stanford, and Rutgers
1983	Improvement of a PDS master framework (CMU)
1983	Improvement of the MED1 master framework (Kaiserslautern)
1984	Clancey and Shortliffe give the accompanying meaning of clinical AI: "Clinical computerized reasoning is worried about the development of AI programs that perform determination and make treatment proposals. Not at all like clinical applications dependent on other programming strategies, for example, have factual and probabilistic techniques, clinical AI programs depended on emblematic models of malady elements and their relationship to tolerant variables and clinical signs" Much has changed from that point forward, and presently, this concept would be viewed as tight in degree and vision. Currently, the significance of finding as an errand need PC assistance in regular clinical circumstances gets substantially less accentuation
1985	Development of MED2 expert system (Kaiserslautern)
Mid-1980s	Foundation of an organization called the fifth Generation Project in Japan, whose goal was to facilitate the communication among medical AI scientists around the world

Mid-1980s	The advancement of master frameworks experienced a couple of downsides as of now: • Some master frameworks could not work just as the specialists who provided them with information • Most of the master frameworks must be run on (expensive) LISP machines • The LISP machines could not be associated with a system These constraints upset the improvement of master frameworks as business applications at that point
1986	Foundation of European Society for Artificial Intelligence in Medicine (AIME) in Europe
1989	Improvement of MUNIN master framework for diagnosing neuromuscular Disarrangement
Early 1990s	The disappointment of scientists to introduce good frameworks brought about diminished subsidizing from the legislature and financial speculators in the mid-1990s. Man-made intelligence researchers regularly allude to the time of 1987–1994 as the “artificial intelligence winter”
1991	Improvement of PEIRS (Pathology Expert Interpretative Reporting System) master framework, to produce pathology reports. PEIRS covered an assortment of pathology measures, with a general indicative exactness of about 95%
1993	Improvement of the GermWatcher AI research center framework. This framework checks for emergency clinic-gained (nosocomial) diseases, which speak to a huge reason for delayed inpatient days and extra medical clinic charges Microbiology culture information from the medical clinic’s laboratory framework is observed by GermWatcher, utilizing a standard base containing a blend of national rules and nearby emergency clinic disease control approach.
Mid-1990s	Research in clinical AI changes its concentration to new territories, including:

	• Getting more information on the web building a better data foundation • Using AI methods to decide • Learning and growing better simple to utilize programs reasonable • for well-being experts A focal piece of every one of these activities was the production of electronic medical records (EMR), which fills in as the focal clinical store of data on persistent consideration
1997	The (American) National Library of Medicine grants agreements to an assortment of human services associations of the nation over to examine inventive employments of the national data framework for social insurance, including telemedicine and data sharing
Near Future	Completing the human genome project will undoubtedly lead to the following AI applications in medicine: • The collection of the information generated through the use of AI systems in hospitals around the world will be linked, enabling them to share all the information. The AI system will aggregate this information utilizing distributed computing and subsequent analysis which will draw inferences from data mining applications that will produce patterns based on this collected data • To determine the critical patterns, expert systems and neural networks will be used • To ensure a system that the system should continue to learn, genetic algorithms should be used The analysis result will be fed back to the AI inference engines of individual hospitals to allow their AI software to analyze data of each patient regarding the present patterns of surgery complications, diseases, medical complications with certain types of the genome, etc.

Appendix B: ***Sample Framework using a Healthcare Provider Example***

Scenario: A healthcare provider wants to implement an algorithm to assist doctors in diagnosing patients with a particular condition.

Workflow Diagram:

1. Data Collection and Analysis:

- Gather patient data, including symptoms, medical history, and test results.
- Analyze the data to identify patterns and correlations using advanced analytics techniques.

2. Bias Identification:

- Scrutinize the data and algorithms for potential biases, such as demographic biases or data collection biases.
- Address and mitigate identified biases to ensure fairness and equity in the diagnostic process.

3. Algorithmic Analysis:

- Develop and refine algorithms based on the analyzed data to improve accuracy and reliability.
- Test the algorithms rigorously to ensure they perform effectively across diverse patient populations.

4. Explainability and Transparency:

- Implement measures to make the algorithm's decision-making process transparent and understandable to healthcare providers.
- Provide insights into how the algorithm arrives at its diagnoses to build trust and facilitate collaboration between the algorithm and healthcare professionals.

5. Customization and Adaptability:

- Design the algorithm to accommodate individual patient variations and evolving medical knowledge.
- Enable customization options to tailor the diagnostic process to specific patient needs and preferences.

6. Functionality:

- Ensure the algorithm can accurately diagnose patients based on input data.
- Verify that the algorithm operates efficiently within the healthcare provider's workflow, minimizing disruption and maximizing productivity.

7. Bias Assessment:

- Periodically assess the algorithm for any emerging biases or unintended consequences.
- Implement mechanisms to address identified biases promptly and effectively.

8. Root Cause Analysis:

- Investigate instances where the algorithm's diagnosis differs from the healthcare provider's assessment.
- Identify underlying reasons for discrepancies and adjust the algorithm accordingly to enhance diagnostic accuracy.

9. Recommendations:

- Provide actionable recommendations based on the algorithm's diagnosis to guide healthcare providers' treatment decisions.
- Offer additional insights or resources to support informed clinical judgment.

10. Monitoring and Feedback Loop:

- Establish a system for continuously monitoring the algorithm's performance and outcomes.
- Solicit feedback from healthcare providers to refine and improve the algorithm's effectiveness and usability continually.

Appendix C: Interview Findings Summary

Theme	*Key Findings*	*Participant Quotes*
Perceptions and understanding of AI Bias	Leaders recognize that AI bias is a significant issue affecting patient care and outcomes.	“Bias is embedded in the data we use; if we don’t address it at the source, we risk perpetuating inequalities.”
	There is a growing awareness of the need for diverse datasets in training algorithms.	“When we talk about human biases, it helps patients see AI as a tool that can potentially offer fairness.”
	Conflicting views exist regarding the severity of AI bias.	“AI is just a tool; it’s how we implement it that matters.”
Impact and Consequences of AI Bias	Biased algorithms can lead to misdiagnoses and inappropriate treatment recommendations.	“We had a case where an algorithm recommended a treatment plan that was not suitable for a patient based on their demographics.”
	AI bias can disrupt workflows and create inefficiencies within healthcare organizations.	“If our systems are directing patients based on biased data, it creates more work for our staff who have to correct those misdirections.”
	Trust in technology is undermined when patients perceive biases in AI systems.	“Patients deserve transparency about how decisions are made; if they feel like they’re being treated by a biased system, we lose their trust.”

Strategies for Mitigating AI Bias	Ensuring diverse and representative datasets is essential for effective AI implementation.	“Building models and collecting data should be representative of the population they’re trying to address.”
	Employing fairness-aware algorithms can help reduce bias from the outset.	“We need to code in fairness from the start; it cannot be an afterthought.”
	Continuous monitoring and auditing of AI systems are critical to identify biases proactively.	“Regular reviews of our algorithms help us identify potential biases before they impact patient care.”
Challenges and Ethical Considerations	Data quality and representativeness are significant barriers to addressing AI bias effectively.	“If we train our algorithms on data that predominantly reflects one demographic, we risk overlooking the unique needs of other groups.”
	Ethical dilemmas arise when balancing technological advancement with equitable care delivery.	“We have to balance innovation with our ethical obligations to provide fair treatment for all patients.”
	Resource limitations hinder comprehensive bias mitigation strategies.	“We want to do more, but resources are tight.”
Communication and Community Engagement	Transparency about AI processes is essential for building trust among patients and providers.	“Transparency is key; if we don’t openly discuss how our algorithms work and their limitations, we risk losing trust.”

	Engaging community stakeholders provides valuable insights into diverse needs and concerns.	"Partnering with local organizations helps us understand community needs better."
	Education campaigns aimed at patients can enhance understanding and acceptance of AI technologies.	"We need our patients to feel informed and empowered about the technology affecting their health."
Future Perspectives and Lessons Learned	Standardized methods for data collection are necessary to mitigate bias effectively.	"We need to establish clear standards for how data is collected and used; otherwise, we will continue to face challenges with biased algorithms."
	Continuous learning opportunities are vital for healthcare professionals regarding AI technologies.	"Training our staff on AI bias is crucial; they need to understand both how to use these tools and their limitations."
	Ethical considerations must remain at the forefront of AI implementation in healthcare.	"We must prioritize ethics over innovation; if we don't do this, we risk harming our patients."

Appendix D: Recruitment by Email

RECRUITMENT BY EMAIL

SUBJECT: INVITATION TO PARTICIPATE IN RESEARCH STUDY ON ALGORITHMIC BIAS IN HEALTHCARE DECISION-MAKING

CONTACT

Email:
kashaw@m.marywood.edu
Phone: 617-448-3921

Dear [Recipient],

I hope this email finds you well. My name is Keishalee Shaw, and I am a researcher at Marywood University. I am currently conducting a study on algorithmic bias in healthcare decision-making, and I would like to invite you to participate.

Your expertise and insights as a healthcare leader would be invaluable to this study. The purpose of the research is to explore the experiences of healthcare leaders in addressing algorithmic bias within the realm of data collection, analysis, and utilization in healthcare decision-making.
Participation in this study involves a confidential interview, which will last approximately 30 minutes. Your responses will be anonymized and treated with the utmost confidentiality. Your involvement will contribute to a better understanding of algorithmic bias in healthcare and help inform strategies to address these challenges.

If you are interested in participating or would like more information about the study, please feel free to contact me at 617-448-3921 or send me an email to kashaw@m.marywood.edu. Your participation would be greatly appreciated, and I look forward to the opportunity to discuss this research with you further.

Thank you for considering this invitation, and I hope to hear from you soon.

Best regards,
Keishalee Shaw
Marywood University

Appendix E: Recruitment by phone

RECRUITMENT BY PHONE

SUBJECT: INVITATION TO PARTICIPATE IN RESEARCH STUDY ON ALGORITHMIC BIAS IN HEALTHCARE DECISION-MAKING

Hello, I hope you're doing well. My name is Keishalee Shaw, and I'm calling from Marywood University. I'm conducting a research study on algorithmic bias in healthcare decision-making, and I'd like to invite you to participate.

Your expertise as a healthcare leader is incredibly valuable for our study. We're interested in hearing from professionals like yourself about your experiences in addressing algorithmic bias within healthcare decision-making.

Participating in our study involves a confidential phone interview, which will take approximately 30 minutes. Your responses will remain confidential and will be treated with the utmost confidentiality. Your insights will play a crucial role in advancing our understanding of algorithmic bias in healthcare and shaping strategies to overcome these challenges.

If you're interested in participating or would like more information about the study, please feel free to call me back at 617-448-3921 or send me an email to kashaw@m.marywood.edu. Your involvement would be greatly appreciated, and I'm eager to discuss this research further with you.

Thank you for considering this opportunity, and I hope to hear from you soon.

Appendix F: Interview Protocol and Qualifying Interview Questions

Qualitative Research Questions

Title of Study: Unveiling Algorithmic Bias: Exploring Healthcare Leaders' Experiences and Strategies for Fairness in AI-Driven Decision-Making"

This qualitative study explores the intricate landscape of algorithmic bias within healthcare, acknowledging the pivotal role of artificial intelligence (AI) and machine learning (ML) algorithms in shaping patient care and resource allocation. Algorithmic bias refers to preferences that arise from designing and implementing AI algorithms, which can lead to consistent mistakes in predicting outcomes. AI involves machines replicating human cognitive abilities by utilizing patterns to match expert systems. ML, a branch of AI, uses data and algorithms to enable AI systems to imitate human learning and improve accuracy over time. In healthcare, biases can occur in AI or ML systems used for diagnosis, resulting in unequal patient outcomes. AI-driven healthcare systems leverage algorithms and data analysis to enhance healthcare delivery for improved efficiency, accuracy, and outcomes.

It builds upon existing literature emphasizing the significance of legal standards, transparency, and fairness in algorithmic decision-making. Drawing on a diverse range of literature, the study aims to understand the experiences of healthcare leaders in addressing algorithmic bias, exploring perceptions, strategies for mitigation, and challenges in ethical considerations. The central research question probes into the lived experiences of healthcare leaders and how these inform strategies for promoting fairness and accountability in AI-driven healthcare systems, with specific sub-questions addressing perception, mitigation strategies, and ethical challenges.

Brief reorientation of the topic: I am interested in learning about your experiences, insights, and recommendations for fostering fairness, transparency, and accountability in healthcare decision-making influenced by AI through this inquiry.

At any time, you feel uncomfortable answering a question, please let me know if you would instead not answer. You can end this interview at any time. Also, if you do not understand a question, please ask for clarification, and I will explain.

Do you have any questions or comments at this point?

Qualifying Interview Questions:

1. In what type of organizational setting do you currently work?
    - Academic/ research/ medical training
    - Biotechnology/ Life sciences
    - Consulting
    - Diagnostics laboratory
    - Government healthcare body
    - Health insurance
    - Hospital/ Practice

- Integrated healthcare delivery system
- Medical equipment manufacturer
- Not For Profit healthcare institution/ professional body
- Pharmaceuticals
- Social services/ care
- Other (open - - to add organization type)

2. What is your professional title at your organization?

3. As a leader specializing in healthcare or health technology, how do you contribute to your organization's utilization of health data and digital technologies for research, diagnosis, and treatment decisions?

4. In your professional capacity, what data types are of primary concern to you?
   - Health/clinical data
   - Operational data
   - Both health-related and operational data
   - Neither health-related nor operational data

5. In your leadership capacity and light of concerns regarding AI algorithmic bias, how would you describe your organization's preparedness for making data-driven decisions at present with the use of Generative AI? For this project, Generative AI is defined as
   - Immature: We utilize some electronic systems or technology for data collection and management, but it is challenging to identify and consolidate relevant data swiftly.
   - Maturing: Most of our data is digital, yet workflow gaps impede rapid diagnosis and decision-making.
   - Mature: We can access, integrate, and analyze data from various sources, enabling prompt and informed decision-making.

6. How many years you have been in the healthcare field?

7. How would you rate your experience in the healthcare field based on this range?

   a. Please rate on a scale of 1-10, with 1 being no experience at all necessary and 10 being the unmatched level of experience.

	1 No experience	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10 Unmatched level of expertise**	**Do not know**
Healthcare Experience											

Appendix G: Interview Questions for Qualified Participants

1. What are your perceptions and beliefs as a healthcare leader regarding algorithmic bias in decision-making within the healthcare industry? Furthermore, how do you perceive the role you play if any in addressing this issue?

2. In your leadership role, considering the potential impacts of AI algorithmic bias, how critical do you perceive it to be for the future of healthcare to effectively oversee the management of data across diverse care settings (such as labs, hospitals, doctors' offices, clinics, patients' homes, etc.)?

3. How do you define algorithmic bias within healthcare decision-making, and what specific examples do they identify?

4. Can you describe instances where algorithmic bias has impacted your organization's patient care or resource allocation?

5. What factors influence healthcare leaders' perceptions of algorithmic bias, such as organizational culture, regulatory environment, or technological capabilities?

6. How do you leaders prioritize addressing algorithmic bias within their organizations, and what motivates their approach to mitigation?

7. What strategies have you used in your organization to identify and mitigate algorithmic bias in healthcare decision-making processes?
    a. Do you think there are other things healthcare leaders or a healthcare leader in your organization implemented to identify and mitigate algorithmic bias in healthcare decision-making processes?

8. How do healthcare leaders collaborate with data scientists, clinicians, and other stakeholders to address algorithmic bias effectively?

9. What challenges do healthcare leaders encounter when implementing bias mitigation strategies, such as data availability, technical limitations, or resistance from within the organization?

10. How do healthcare leaders balance the need for algorithmic accuracy with considerations of fairness, equity, and patient outcomes?

11. What ethical considerations do healthcare leaders grapple with when addressing algorithmic bias, particularly regarding patient privacy, consent, and autonomy?

12. How do healthcare leaders ensure transparency and accountability in AI-driven healthcare systems, especially concerning algorithmic decision-making?

13. Can you share examples of successful interventions or initiatives within your organization that aim to reduce algorithmic bias in healthcare decision-making?

14. How do healthcare leaders evaluate the effectiveness of bias mitigation strategies, and what metrics or indicators do they use?

15. How do regulatory standards and guidelines shape healthcare leaders' approach to addressing algorithmic bias?

16. How do healthcare leaders promote diversity and inclusivity in algorithm development teams to mitigate biases rooted in data collection and analysis?

17. What resources or support do healthcare leaders need to enhance their capacity to address algorithmic bias effectively?

18. How do healthcare leaders communicate with patients and the broader community about algorithmic decision-making in healthcare, particularly concerning transparency and trust?

19. How do healthcare leaders navigate tensions between innovation and risk mitigation when deploying AI-driven solutions in healthcare settings?

20. Can you describe any instances where healthcare leaders have had to make difficult decisions regarding algorithmic bias, and how did they approach these challenges?

21. What lessons have healthcare leaders learned from previous experiences in addressing algorithmic bias, and how have these lessons informed their current strategies?

22. In what ways do healthcare leaders envision the future of AI-driven healthcare systems, and how do they plan to ensure fairness, transparency, and accountability in this evolving landscape?

23. Is there anything else you want to share about Healthcare Leaders' Experiences and Strategies for Fairness in AI-Driven Decision-Making?