Investigating the Complexities and Interdependencies of Algorithmic Biases in Healthcare Artificial Intelligence

Ruth Bahta

Candidate for Master of Public Policy and Administration

California State University, Sacramento

May 2024

Advisor: Ahrum Chang, Ph.D.

Table of Contents

Executive Summary

The current challenge in implementing Artificial Intelligence (AI) in healthcare is algorithmic biases resulting in racial biases. Inequities in diagnosis, treatment, and billing have increased, disproportionality impacting people of color. This study aims to understand diverse problem framings on algorithmic biases from patients, healthcare professionals, AI developers/technology companies, and policy-making bodies. In particular, it provides takeaways for educating and empowering patients and healthcare professionals in advocacy efforts. The research is conducted within the lens of wicked problems, where stakeholders agree on the problem's existence but differ in their problem framing, thus delaying mitigation plans. The findings are assessed through the multiple-stream framework, enhancing understanding of policy dynamics through three independent streams: problem, policy, and politics, which are only merged by a policy entrepreneur for policy change to occur.

The findings illustrate multifaceted problems stemming from homogenous datasets, lack of supportive guidelines for clinicians to use along with the technology, underrepresentation in the AI developer's workforce, and lack of regulations in the algorithmic life cycle. Furthermore, the analysis indicates an overarching theme of underrepresentation in training data due to historically existing bias embedment while exhibiting additional problem framing unique to the stakeholder's context. The study provides recommendations for further research and engagement takeaways for patients and healthcare professionals, centered on enhancing understanding of policy dynamics, opportunities to get involved in current initiatives, and strategies to empower underserved communities.

Introduction

History and Background of Artificial Intelligence in Healthcare

The 1960s marked the birth of Artificial Intelligence (AI) in healthcare. Starting from the Dentral Project at Stanford University, analyzing mass spectrometry data for chemical analysis, laying the groundwork for future medical applications (Jiang F, Jiang Y, Zhi, et al., 2017). Then, MYCIN, an early expert system created at Stanford University in the 1970s, demonstrated the promise of AI in medical decision-making by identifying bacterial illnesses and recommending antibiotic therapies (Copeland, 2018). Machine learning in healthcare began to gain traction in the 2000s, when image recognition breakthroughs, personalized medicine, genomics, AI in drug discovery, telehealth, and AI-assisted diagnostics transformed the healthcare delivery system.

As integrating various technology tools enhanced patient outcomes, their unintended consequences resulted in patient safety, data privacy, and algorithmic biases issues. In particular, algorithmic biases gained recognition in the late 2010s due to several high-profile studies and incidents. Algorithmic biases are "when the application of an algorithm compounds existing inequities in socioeconomic status, race, ethnic background, religion, gender, disability or sexual orientation to amplify them and adversely impact inequities in health systems" (Panch et al., 2019). A prominent illustration is when machine learning techniques were used to detect skin cancer via skin image datasets. Wen et al. (2022) concluded that the datasets predominantly collected images of lighter skin types. Thus, the algorithm-based diagnosis detected skin cancer on the lighter skin more, underreporting for darker skin.

Similarly, Obermeyer et al. (2019) and Rajkumar et al. (2018) study highlight the racial disparities in healthcare algorithms through predictive modeling and diagnostic accuracy tests, resulting in three main impacts on patients, healthcare providers, and policymakers. In particular,

Obermeyer et al. (2019) discovered systematic discrimination against black patients when hospitals and insurers used commercial risk-prediction algorithms, which allocated medical care to thousands of patients predominantly white. The algorithm drew the yearly healthcare costs from electronic health records to assess the healthcare needs. The calculation indicated that white patients had higher risk scores needing to receive more personalized care, while black needing less care. The authors concluded that there are social-political-economic factors impacting the results to be biased as black patients deal with economic barriers, racial discrimination, and historical distrust in the healthcare system. All signifying that across all stages of healthcare delivery, racial and ethnic, gender, and socioeconomic biases are prevalent.

The Importance of Addressing Algorithmic Biases

Healthcare algorithms are the assortment and process of medical records, clinical drug trials, insurance claims, and other sources. Studies such as in skin cancer detection, gender bias in heart disease, and language barriers are cases contributing to the algorithmic bias. Historical cognitive biases are influencing AI, contributing to the health disparities for people of color. The current challenge in implementing AI in healthcare is algorithmic biases resulting in racial biases. In return, inequities in diagnosis, treatment, and billing have increased.

This study investigates the complexities and interdependencies of algorithmic biases in AI within healthcare. In particular, it explores the wicked problem, where there is consensus among stakeholders on the problem's existence, but all have different problem framing. Straus (2010) points out how clear and measurable identification of perception, definition, and analysis are the building blocks to bringing consensus among stakeholders. Common issues delaying collaborative work among stakeholders are ambiguity and unattainable goals, resulting in ineffective communication and wasted resources. Thus, understanding the problem framing from

relevant stakeholders is crucial in navigating complex collaborations that require agreements. Incorporating the multiple stream frameworks will further the analysis, providing clarity and guidance to enhance transparency and effectiveness in leading collaborative efforts for patients and healthcare professionals.

Literature Review

Wicked Problem

Rittel and Webber (1973) introduced wicked problems in their seminal paper, signifying the complexity and resistance of its nature to traditional problem-solving techniques. Although there is an agreement on an issue, the framing of the problem from all stakeholders involved makes it difficult to come to a consensus. Thus, alternatives can be generated, but due to the complexities and interdependencies of the problem framing, intergenerational problems persist today, shaping modern society. Westcombe (2007) highlights the importance of building shared understanding to build a foundation among stakeholders in addressing complex issues. Head (2022) demonstrates the application of wicked problems to a public policy to generate flexible and innovative solutions. The use of AI in healthcare settings is relatively new research. The development, implementation, and impact are best captured through understanding relevant stakeholders: patients, health professionals, AI developers/technology companies, and policy-making bodies. Although additional stakeholders exist, these four are the key players shaping the healthcare AI landscape.

Stakeholders Problem Framings

Patients. Patients are individuals receiving medical services, diagnosis, and treatment. They come from a diverse demographic, genders, ages, ethnicities, races, and socioeconomic statuses. Their participation is a wide range, as users of health-related AI applications,

participants in clinical research, or/and a care recipient. As the primary beneficiary and target audience for healthcare AI devices, they are the central stakeholders in the discussion around algorithmic biases.

In their study, Richardson et al. (2021) explore the apprehension surrounding the use of AI in healthcare. There is excitement about implementing new technology, but significant patient concerns exist regarding safety, patient autonomy, data integrity, and reliance on technology. All concerns highlight the importance of physician oversight and patient autonomy to mitigate data biases. They advocate that clinicians should have discretion over treatment plans to buffer the potential harm of the accuracy and reliability of data used to train AI systems. They indicate that as a shared data source for AI, Electronic Health Records (EHRs) containing omissions or errors can inaccurately reflect patients' information during diagnosis and treatment. They also expressed the possibility of AI tools perpetuating existing biases due to the homogeneous datasets leading to technological failures.

Furthermore, Straw (2020) highlights that demographic health inequities continue to impact medical care as technology integration increases. He argues that historically, the population of professionals designing the field is dominated by a narrow demographic group. This alters the needs, perspectives, and understanding of the issue. For instance, there has been a lack of racial/ethnic, women, and gender minorities in the AI workforce in the last century. He utilizes an example from the "Weapons of Math Destruction," highlighting that rather than actual crime data being used, arrest data are used to predict policing algorithms. Historical, racial/ethnic minorities and low-income groups are disproportionately arrested at a higher rate compared to their higher-income counterparts. Thus, discriminatory practices embedded in the datasets are reflected more than the actual crime rate.

Similarly, algorithmic-based decisions stem from an already flawed database. In medicine, implicit biases from non-representative physicians can result in the embedding of a flawed diagnosis rather than the actual disease rates, widening health disparities. Therefore, Richardson et al. (2021) and Straw (2020) indicate patients as the primary beneficiaries of medical care, pointing out a lack of diversity in training data and healthcare professionals' discretion in diagnosis and treatment plans. The overarching theme delves into the lack of transparency and accountability in AI development and implementation.

Healthcare Professionals. They play critical roles in clinical care and healthcare administration. Depending on their position, they encompass various responsibilities and activities, such as diagnosis, treatment plans, and medical care administration. They contribute to the broader goals of improving patient outcomes, public health, and healthcare delivery. They are not limited to patient education, care, and administration and are essential contributors to research and policy. They are uniquely positioned to identify and recognize biases directly by working with patients and bridging clinical practice and AI development through ethical oversight.

O'Connor & Booth (2022) underline that healthcare professionals have therapeutic relationships with patients but need more awareness of the risks associated with AI technologies and techniques, particularly in algorithmic biases. Nurses and healthcare professionals are generally prone to burnout due to the nature of the healthcare job. Therefore, the likelihood of allocating time and effort to learning about the technologies without guidelines could be higher. Thus, the authors advocate for nurses to enroll in a curriculum covering the "fundamentals of AI computational techniques and the ethics issues it brings, such as algorithmic bias, need to be included in undergraduate and postgraduate programs to educate nurse students and the

workforce" (Booth, Strudwick, McBride, O'Connor, & Lopez, 2021). In addition, he believes that involvement in legislative and policy changes can shape the development and governance of AI in healthcare. All of these highlight the lack of investment in healthcare professionals' education and legislative regulations to address issues with AI-based technologies implemented for medical care.

Likewise, Aquino's (2023) study discusses existing clinical and social biases replicating algorithmic biases in healthcare in Figure 1. The dataset is the central source for compiling human input from medical records. These records express the incorporation of existing unchecked biases in the real world.

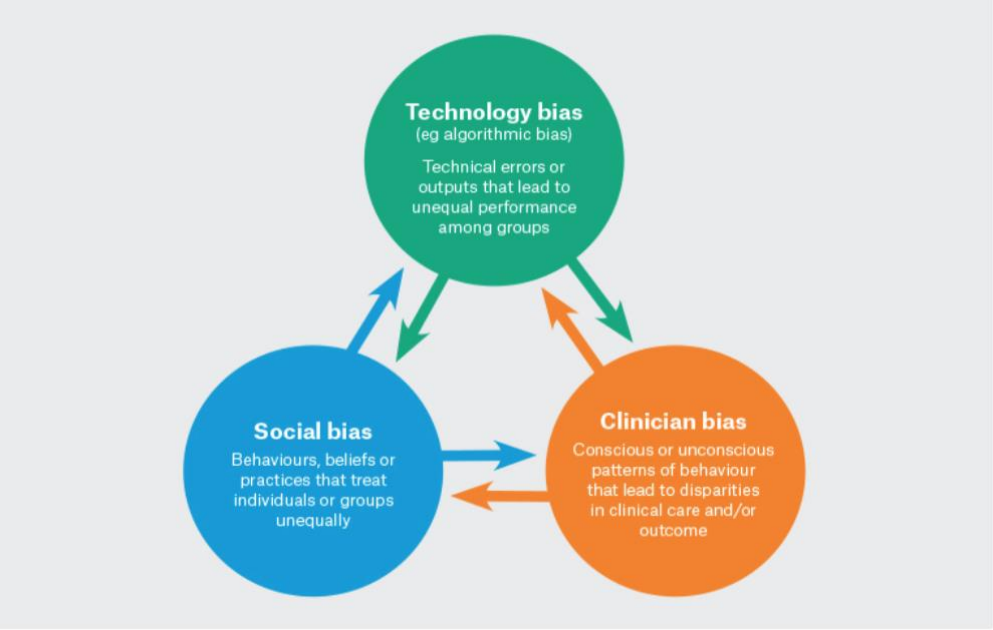

Figure 1: Cycle of Bias (Aquino, 2023)

As a result, patient safety and care outcomes are altered, leading to unequal treatment for individuals or groups. For instance, pulse oximeters use light to measure blood oxygen, but found that the reading accuracy on darker skin tones decreases. Demographic variability in diagnostics and testing highlights the systematic issues of the technology, disproportionately

affecting people of color. Also, he indicates that gender-based disparities in cardiac event prediction predominantly trained on male patient datasets result in underdiagnosis or misdiagnosis for female patients. Scholars underscore the lack of education for clinicians and diversity in AI training datasets, including gender-balanced data. The technical challenges perpetuate disparities for underserved communities and women, and broader social inequalities exist (Aquino, 2023; O'Connor & Booth, 2022).

AI Developers/Technology Companies. They birth the technology from research, design, and development to implementation. They pioneer various technologies to address healthcare needs, ensuring reliability, accuracy, and ethical use. They work with healthcare professionals and regulatory entities to navigate data privacy and security. Hague (2019) underscores the complexity of algorithmic biases, impacting care, claims, and marketing in recognition that they lead to unequal treatment and outcomes. He explores frameworks used in other industries regarding data management, machine, and human biases arising from the implementation of AI. A central theme is drawn from various studies, such as the genetic diseases risk-detection algorithm and Go Red for Women by the American Health Association. Illustrative through the examining claim/payment processing model, a hypothetical scenario. The case study pointed out the model approving claims for wealthier and white individuals, possessing the highest-quality health plans in comparison to their counterparts, ethnic minorities. Also, the marketing approach is driven by higher return investment; thus, the model uses a dataset lacking a diverse patient pool to generate strategies that benefit those who can pay. This finding indicated historical data perpetuating exciting inequalities as a critical factor contributing to algorithmic biases. The author concluded that AI developers and technology companies

attribute algorithmic bias challenges to a lack of diversity in training datasets and consideration of additional ethics.

In comparison, Choudhury and Asan (2020) explore perspectives addressing algorithmic biases by highlighting Human Factors and Ergonomics (HFE). Developers recognize patient safety as the central issue of misinterpretation and poor utilization by clinicians due to the complexity of AI technologies. The author argues that the opacity and complexity of AI complicate the output utilization by healthcare professionals, leading to the risk of misinterpretations. Thus, it advocates the integration of the HFE in the design process and simplifies AI interference. In return, it optimizes human performance and clinicians' workflow. Likewise, Lamanna and Byrne (2018) argue that the historical datasets from electronic health records reflect and propagate existing biases in the healthcare system. They propose the "autonomy algorithm" to help elderly and psychiatric patients who cannot make informed healthcare decisions. This model incorporates diverse datasets, including demographics, records of healthcare interventions, and social media behaviors. Although it mitigates existing methods such as surrogate decision-making, it highlights the importance of healthcare professionals' critical evaluation and advocates for utilizing the technology as an aiding tool. Based on the various arguments, the problem framing is multifaceted, focusing on the lack of diversity in training datasets and human-centered designs.

Policy-Making Bodies. Policymakers are relatively new to AI integration in the workforce, particularly in healthcare. Even state agencies such as California Healthcare of Access and Information have not fully addressed AI. However, various entities such as the World Health Organization advocate through creating awareness around the subject. In addition, recent research explores legislative responses to the impact of AI in the healthcare delivery

system. In their review, Nazer et al. (2023) present the multifaceted issue stemming from several stages: problem formulation, data collection, preprocessing, development, validation, and full implementation. For instance, the authors explore biases in direct model, data collection, and preprocessing. The direct model bias demonstrates the inadvertent favoritism of healthier white patients obtaining care over sicker black patients due to the cost prediction on healthcare needs. The bias in data collection and preprocessing spotlights the lack of data reliability, especially in the development of algorithm prediction for Acute Kidney Injury (AKI) by the US Department of Veteran Affairs. When assessed, the data collection on this injury predominantly focused on older white men. Thus affecting ethnic minorities disproportionately, leading to misdiagnosis or underdiagnosis. In particular, the legislative perspective emphasizes the lack of inclusion of diverse viewpoints on the problem framing at the development stage.

Similarly, Thais et al. (2023) argue that the entire algorithmic life cycle in Figure 2 and the statistical biases in training data contribute to the algorithmic biases in healthcare AI.

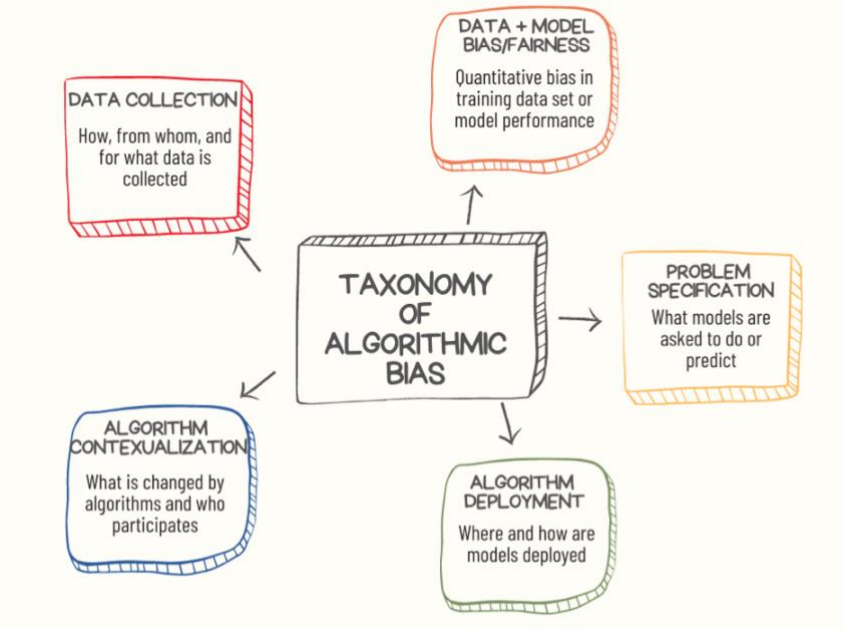

Figure 2: Taxonomy of Algorithmic Bias (Thais et al., 2023)

The data collection, problem definition, determination of AI use, and the broader social issues at play make up the entire life cycle. Illustrative through the Allegheny Family Screening Tool (AFST) application in Allegheny County, Pennsylvania, where black families were targeted for investigation due to technical errors. The AFST aims to help in child welfare decisions to protect vulnerable children. There was positive expectation and hope that the technology would foster safer communities by identifying at-risk families early on. After implantation, a higher proportion of black families were targeted for investigation, highlighting bias and fairness in the decision-making process. Assessing the issues, historical data, and social disparities were attributes in the datasets the algorithms used to make decisions. The decision stemmed from existing inequality rather than intentional prejudice from its creator. This sparked community-wide decisions about fairness, equity, and justice. The experience underscored the importance of continuous improvement of AI datasets in the public sector. It advocated for critical examination and diverse AI system input mitigating societal inequalities.

This case study is a stepping stone to balancing innovation with fairness, equity, and justice. Moreover, the research highlights the gap in policy approach due to a narrowed focus on quantitative fairness, undermining the sociotechnical complexities of AI implementation. Therefore, advocate for legislative and regulatory entities to create an inclusive and adaptive framework mitigating the evolving landscapes of technology and societal needs.

Overview of Problem Framings

In similarity, patients, healthcare professionals, AI developers and technology companies, and policy-making bodies all agree there needs to be more representation in the training data, undermining the socio-technical complexities. It reflects the integration of existing unchecked biases in the world. Second, excluding the policy-making bodies, the rest of the stakeholders

agree there needs to be more clinician discretion stems from the limited awareness and education on the risk associated with the technology. Thus, the technology exhibits less human-centered design due to the need for more diverse professionals to design the AI program or tool. A common theme centered around the policy-making bodies is the recognition of multifaceted root problems at every stage of the algorithmic life cycle, particularly the lack of diverse viewpoints at the development stage. Therefore, it highlights the need for a more innovative legislative and regulatory framework. All stakeholders discussed above highlight diverse problem framings, exhibiting similarities and differences. This illustrates the complexities stakeholders face in obtaining consensus on the problem framing to mitigate issues promptly.

Moreover, the current frameworks applied in the realms of algorithmic biases in healthcare AI encompass various strategies and frameworks. Most notably the following:

- **Comprehensive Review and Mitigation Strategies:** Identify potential sources of biases at the development and implementation stage of healthcare AI algorithms to establish equitable health outcomes (Nazer et al., 2023).
- **Closed Loop Framework:** Assesses bias by evaluating data representativeness, feature bias, and subgroup validity and integrates fairness criteria to tailor the care management enrolment threshold (McCall et al., 2022).
- **Explainable AI (XAI):** Aims to make AI decisions understandable, support ethical use in patient care, and enhance transparency and accountability in AI (Upadhyay et al., 2023).
- **Human-Centered Design:** A form of a framework that underscores diverse stakeholders' involvement in all stages of the AI lifecycle to reduce health disparities (Chen et al., 2023).

- **Total Product Lifecycle Framework:** Elaborate framework mitigating biases across AI systems' entire lifecycle, thus enhancing patient health outcomes (Abramoff et al., 2023).

These collective frameworks and designs are meant to tackle algorithmic biases in development and implementation. Their application in various research aims to improve overall health outcomes in healthcare AI. However, they must be more comprehensive to keep up with the evolving technological changes.

Methods

This study investigates the complexities of algorithm biases in healthcare AI. To understand the multifaceted problem, relevant stakeholders—patients, healthcare professionals, AI developers/technology companies, and policy-making bodies—perspectives are considered. This methodology serves as a map of awareness, education, and empowerment for patients and healthcare professionals directly impacted by this issue.

The research field surrounding healthcare AI is limited compared to other health-related topics. Therefore, a qualitative study is best suited for tackling a wicked problem, enhancing understanding of diverse stakeholders' perspectives, and the nuances of bias in AI. The scope of the study centers on the healthcare and technology industry within the United States but is not limited globally.

Since this study has a time constraint, primary sources such as interviews from relevant stakeholders are not used. However, secondary sources from peer-reviewed articles, case studies, and policy documentation on healthcare AI are incorporated. The criteria to evaluate the sources anchor in finding authors and informants representing the stakeholders such as medical doctors and researchers to represent the healthcare professionals' perspective. Also, the types of biases sourced from the research are data biases and model biases. The data bias pertains to the

diversity of the dataset used to train an AI model. It anchors in the underrepresentation or overrepresentation of diverse patients due to existing historical biases in the real-world environment.

In comparison, the model bias focuses on the AI development phases with the introduction assumptions, structure, or learning algorithm. For instance, a trained model can assume and overlook the socioeconomic factors affecting healthcare access. In this case, initial biases are reinforced when a model learns from its outputs by developing a feedback loop, resulting in misinterpretations, disproportionately affecting underserved communities. The reliability and findings of the sources are from credible research databases with peer-reviewed studies and audited governmental sources.

The Multiple Stream Framework (MSF) explores the nature of how policy change occurs. In the 1980s, John W. Kingdon, a political scientist, introduced MSF to explain the formulation of policies in the public sector and capture the change with the interdependencies of three streams: problems, policies, and politics. A public policy agenda is formulated based on evaluations of agenda-setting procedures inside the fractured government political system. The three streams are independent and interrelated variables that work together to provide "windows of opportunity" for agenda setting. Until a policy window opens at a specific time, these three streams travel different paths and operate mostly independently. Just then, as the streams meet, the change can occur. It is also possible that external focus events with unrelated connections, such as crises, accidents, or the existence or absence of "policy entrepreneurs" both inside and outside of governments, might cause the windows to open.

Moreover, the MSF nurtures educational content to equip patients and healthcare professionals with the knowledge, skills, and techniques to mitigate this issue. In addition, it

advocates for innovative approaches in tackling algorithmic biases in healthcare, optimizing patient safety and healthcare professional productivity. Therefore, the research design, data sources, selection criteria, framework, and operational definitions align with this study's aim of understanding perspectives from diverse stakeholders regarding algorithmic biases in healthcare AI.

Analysis and Findings

Current implemented frameworks do not map out how the different problem framing of relevant stakeholders interacts within policy dynamics such as the MSF. This framework will serve as a lens to analyze problem framing within the scope of the following three streams.

Problem Stream

The problem centers around the consensus among relevant stakeholders recognizing the existence of algorithmic biases in healthcare AI. However, stakeholders differ in problem framing, leading to prolonged mitigation processes. As the literature review identifies, the problem has alignments and contrasting framing. The patient's perspective exhibits mistrust and fear of delayed diagnosis or misdiagnosis due to underrepresentation in training data and lack of patient and clinician autonomy. Healthcare professionals need more understanding and training on using these technologies to make informed decisions. The embedding of historical existing biases in the datasets and lack of human-centered design challenges AI developers/technology companies. At last, policy-making bodies lack diverse viewpoints and regulations in the development stage and algorithmic lifecycle and underscore statistical biases in training data.

Although different factors contribute to the issue, there is a general alignment with differing angles focusing on the lack of representation in the training data overlooking the socio-technical complexities. Interestingly, Vorisek et al. (2023) surveyed AI developers assessing the

perception of AI biases in healthcare. As a result, one-third of the AI developers rated their AI projects as moderately fair to fair, while a minority rated them as not fair to somewhat fair. The demographic breakdown underscores the need for more focused study on ethnic minorities and women as they perceive AI biases in healthcare to be negatively impacting underrepresented communities. Moreover, the survey revealed the differing data sources and types, where half of the participants worked on image data from single-center sources, while the majority worked on national data. The lack of data sources and types are potential limitations in capturing diverse and representative training datasets. Not only is there a lack of representation in the training data, but also a lack of representation in the AI developer/technology companies' workforce. Overall, there is consensus regarding the prevalence of biases. However, it is attributable to either a lack of general knowledge of the biases, inadequate guidelines, or a lack of fair data, indicating a gap in the systematic approach.

Policy Stream

This stream collects policy alternatives and proposals from various stakeholders, such as interest groups, academics, and experts. In this phase, the options are generated and refined. Policies fall under two categories: general and precise regulation. The general regulation applies across different sectors, while precise regulations are explicitly tailored to an industry such as healthcare. According to Reddy et al, (2023), the following are an example of the current regulations landscape in the United States:

- **FDA's AI/Machine Learning-Based Software as a Medical Device Action Plan:** Aims to ensure safety and effectiveness over AI/Machine learning software life cycle by establishing a framework for the continual assessment and validation of AI applications in healthcare

- **National Institute of Standards and Technology (NIST) in the U.S.A.:** NIST ensures patient safety and data privacy by developing guiding standards for AI implementation in healthcare.

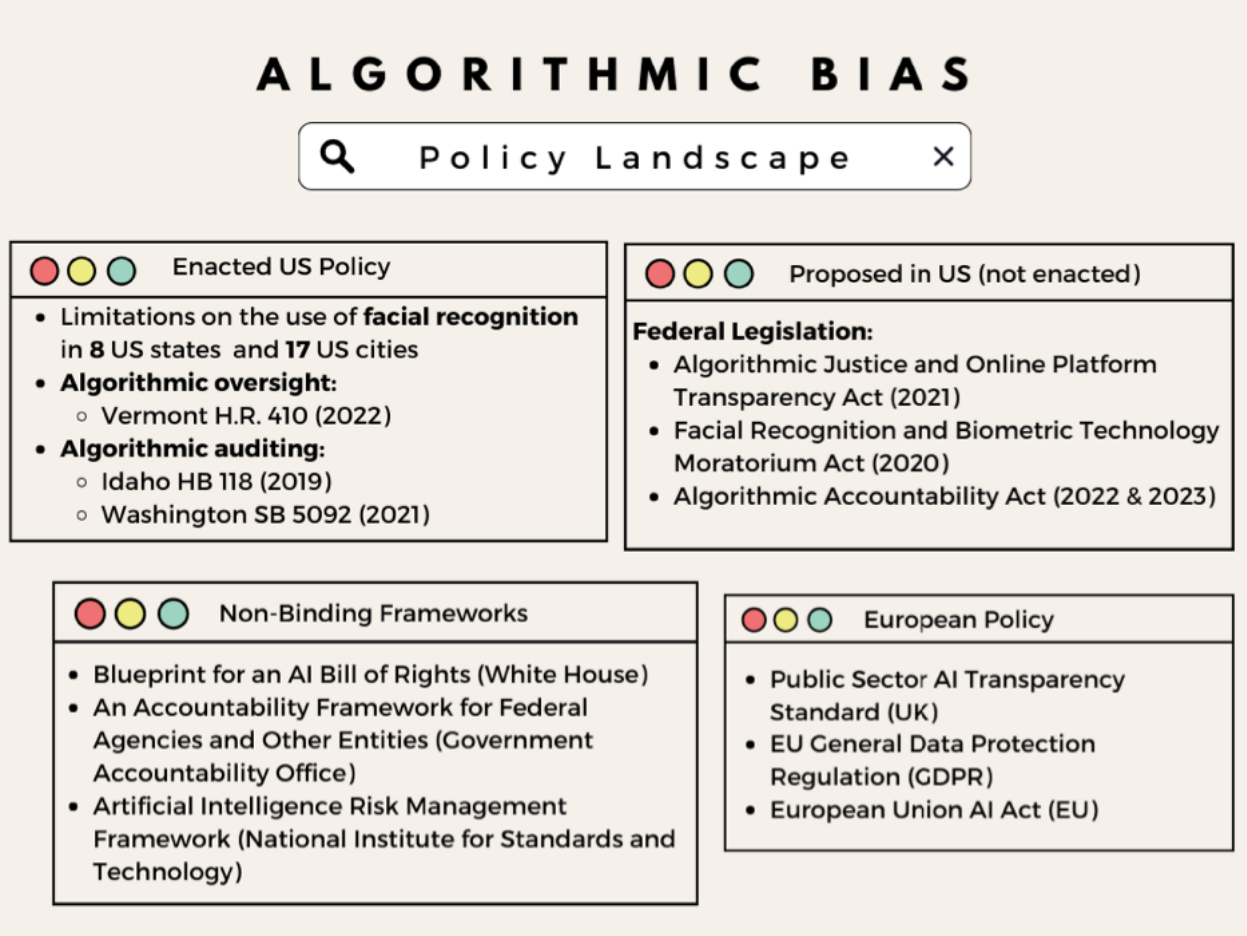

Figure 3: Summary of Algorithmic Biases in the Policy Landscape (Thais et al., 2023)

According to Figure 3, the scholars capture the current US and global policy in four categories: enacted US policy, proposed in the US (not enacted), non-binding frameworks, and European policy. There is a focus on temporary bans for certain government bodies at the local and state levels targeting facial recognition technologies to address privacy and bias concerns. In Vermont, HA 410 requires an inventory and public disclosure for automated decision-making systems to address accountability and transparency in the public sector. Similarly, Washington's

SB 5092 provides funding for a workgroup to research how automatic decision-making systems can be reviewed and audited.

In 2022, California Attorney General Rob Bonta asked 30 hospital CEOs to report detailed information about implementing and overseeing healthcare decision-making tools (Bonta, 2022) to ensure equitable access to quality healthcare. This investigation investigated how these tools impact and mitigate the ethical disparities surrounding algorithms. These inquiries serve as the ongoing compliance and assessment of equity.

Moreover, Assemblymember Rebecca Bauer-Kahan introduced AB 2930, a bill addressing algorithmic biases generated from automated tools across sectors, such as healthcare, housing, and employment. It requires the developers and Automated Decision Tools (ADTs) users to conduct an impact assessment on the intended use, the makeup of the data, and statistical analysis (Bauer-Kahan, 2024). The bills stem from understanding how ADTs are integrated within insurance eligibility, employment screening, credit decisions, and healthcare to determine benefits and penalty eligibility factors. Through studies, this integration has been shown to produce biases resulting in discrimination harming underserved communities. This bill mandates public disclosure of AI-based decision-making systems to enhance government and public trust. The bill is re-referred to the Committee on Appropriation, which is read the second time and amended as of April 24, 2024. This legislation is based on AB 331, the White House Blueprint for Artificial Intelligence, and the National Institute of Standards and Technology framework.

An additional innovative policy proposed by the Office of National Coordinator for Health Information Technology is the Health Data, Technology, and Interoperability: Certification Program Updates, Algorithm Transparency, and Information Sharing (HTI-1) Final

Rule addresses AI applications in healthcare and health IT to enhance transparency and interoperability (Office of the National Coordinator for Health Information Technology, 2024). Critical Aspects of the HTI-1 Final Rule:

- **Algorithm Transparency:** Establishes transparency criteria for AI and other predictive algorithms. It ensures clinicians access consistent information when assessing the tools for fairness, validity, appropriateness, effectiveness, and safety.
- **Adoptive of USCDI Version 3:** Establishes a new baseline standard with the United States Core Data for Interoperability (USCIS) and the ONC Health IT Certification Program starting January 1, 2026. This ensures more accurate and complete patient characteristics data updates to promote equity.
- **Enhanced Information Blocking Requirements:** This policy utilizes the Trust Exchange Framework and Common Agreement (TEFCA) to revise definitions and exceptions to information blocking to support electronic health information sharing. It promotes efficient, secure, and standard-based information encouragement.
- **New Interoperability-focused reporting metrics for Certified Health IT:** Require metric-based reports on developers' participation in certification programs in health IT.

These key aspects became effective on March 11, 2024, as a step to improving oversight and regulations in publicizing developers' risk management practices in alignment with federal initiatives to foster equity and safety.

Overall, current legislation has yet to be enacted nationally to address algorithmic biases in healthcare. Only such as White House's Blueprint for an AI Bill of Rights indirectly influences regulations addressing AI biases in general and listed legislation in Figure 3. However, California spearheaded the California Interagency AI Working Group (SB 721) and adapted

Senate Concurrent Resolution No. 17 (SCR 17). SB 721 is currently pending as it proposes establishing a task force to investigate the application of AI in healthcare (Becker, 2023). SCR 17 is a resolution in compliance with President Biden's vision for safe AI use and principles outlined in the Blueprint for an AI for an AI Bill of Rights (Dodd, 2023). The pandemic has catalyzed the integration of technologies in the healthcare setting. Thus, it encourages involved stakeholders to pay more attention to finding ways to mitigate issues surrounding patient safety, data privacy, and algorithmic biases. California has been a trendsetter in the public and private sectors in laying the guardrails to protect citizens from the consequential decisions driven by AI in healthcare.

Politics Stream

The political stream assesses the political climate, such as a change in administration, partisanship, and national mood, to evaluate alternatives' political feasibility. Change in the federal administration influences funding for priority projects based on the policy agenda and regulatory approaches. The change from Trump to Biden's administration resulted in a focus on tackling algorithm biases across different sectors. Key examples are:

- **Biden's Executive Order on Trustworthy AI:** This executive order was initiated to create a guideline for federal agencies to develop targeted standards, emphasizing AI regulation benefiting end users. This systematic approach ensures that AI technologies are safe, secure, and beneficial for the public (The White House, 2023).
- **The White House AI Bill of Rights:** A broad outline of ethical principles to ensure fairness, transparency, and accountability during the development and implementation of AI (Bauer-Kahan, 2024).

- **Voluntary Commitments by Healthcare Organizations:** In alignment with the Biden-Harris administration, 28 healthcare providers and payer organizations established a mutual agreement to develop AI solutions to optimize healthcare delivery and payment systems by adhering to principles of fairness and safety in the use of AI tools (U.S Department of Health and Human Services, 2023).

From the partisanship angle, Democrats advocate for stricter technology regulation, while Republicans promote lighter regulators prioritizing innovation and economic benefits. For instance, the House Energy and Commerce Committee Chair, Cathy McMorris Rodgers (R-Wash), advocated for a broader-based private law to regulate AI in healthcare and beyond (Rodger, 2023). In particular, it emphasizes less excessive regulation in not burdening technological innovation but focusing more on creating foundational privacy protections. In contrast, Democrats tend to advocate for stricter regulation to address biases in AI with a cautious approach. Partisanship influences the legislative process and the nature of regulation. Therefore, comprehensive bipartisanship support is vital to addressing the healthcare system's AI complexities.

Moreover, policy responses are influenced by national mood and public opinion. High-profile incidents of biases in healthcare are crucial to creating collaborative advocacy to address algorithm biases in healthcare. For example, the COVID-19 pandemic is a pinnacle, resulting in increased public awareness influencing regulator actions and legislative priorities. The rapid implementation of AI tools during the pandemic highlights the ethical considerations and potential risks with AI. In particular, the California Department of Public Health (CDPH) “Let's Get Health California " initiative provided information to California regarding updates to health services and updated health guidelines (California Department of Public Health, 2014). This led

to a significant increase in public calls and engagement with health-related legislative processes. Therefore, increasing public awareness of algorithm bias impacts can lead to significant demand for regulatory oversight.

Overall, the role of the policy entrepreneur is critical to evaluating the political feasibility. This entrepreneur can be an individual or group actively preparing and leveraging resources to create urgency for policy change. This requires the coordination of collaborative partnership, strategic advocacy, and transparent communication for public engagement via a window of opportunity to spearhead a sustainable policy change.

Discussion

Based on the findings and analysis, the challenges in addressing algorithm biases are a need for more diverse data, continual monitoring, and ethical and legal regulations. Each stakeholder has a critical role in collaborating to bring change in the policy realm. This research focuses, in particular, on informing patients and healthcare professionals on how to strategize for engagement with relevant stakeholders to ensure a comprehensive and holistic approach to mitigate the issue. The following three takeaways are foundational in addressing algorithmic biases in healthcare AI with expected outcomes and potential barriers.

Takeaways

First, focus on learning more about policy dynamics. In particular, alongside MSF, Punctuated Equilibrium Theory (PET) is complementary in equipping patients and healthcare professionals to strategize their advocacy efforts in addressing algorithm biases. The idea of PET measures and explains long periods of status quo or stasis in policy are interspersed by quick but extraordinarily intense times of instability and upheaval. Policymaking quickens and then slows down when worries appear on the public agenda and disappear. Policy pictures are crucial for

moving themes outside the purview of particular interests and expert organizations that could monopolize policy. There are two systems in operation. The macro-political system includes Congress and elected officials. The policy subsystems consist of interest organizations and policy entrepreneurs. The macro system is designed to maintain the status quo and display constrained rationality, allowing public employees to concentrate on just a portion of the issues that fall within their jurisdiction. Because of policy monopolies' unfavorable feedback, the status quo is present. Positive feedback from the subsystems helps to break the status quo. The intervention begins when venue shopping or window of opportunity matches knowledge about the intervention's issue with a macro-level focused event. MSF and PET can serve as a guiding framework and theory to understand roles and policy dynamics diffusing potential resistance as they provide an informative resource.

Second, patients and healthcare professionals can advocate by collaborating with groups, organizations, and applicable entities at the local or state level to work on mitigating algorithmic biases. First, *A Citizen's Guide to Participation* is an informative guide to learning how to engage with the legislative process as a citizen. Second, alongside the bill discussed in this research, there are vital bills to support and get involved in according to California Legislative Information (2024):

- **AB 85 (Weber) Social Determinants of Health: Screening and Outreach:** Requires health plans and insurers to pay for the screening for social determinants of health. Although vetoed, supporting similar bills that can equip relevant stakeholders such as AI developers and healthcare teams with diversified representation datasets, encompassing the socio-technical complexities is key.

- **AB 2058 (Weber) Automated Decision Systems:** Requires a comprehensive inventory of all high-risk automated decision systems in commercial algorithms and AI-enabled medical devices.
- **AB 331 (Bauer-Kahan) Automated Decision Tools:** Aims to establish standards for evaluating the impacts of automated decision tools and regulating algorithmic biases.

Additional bills can be accessed through the Legislative Information websites. As a result, supporting existing bills provides the initial step without having to start from scratch to learn and engage in evidence-based initiatives. This opens networking opportunities with allies to find related projects to contribute to and make a meaningful difference. Some oppositions work against the bills because interest groups benefit from the status quo by funding lobbyists. Also, the current California budget deficit can impact project funding, primarily when a monetary ask exists. It might not be necessary due to opposition, but competing bills can result in delaying or vetoing the bill.

Third, underserved communities are often impacted by the healthcare system. Therefore, implementing the Nudge theory as a strategy uses a cognitive bias to encourage community members toward the desired behaviors without limiting freedom of choice. For instance, strategize to reach communities by understanding what is important to them. Collaborating with leaders respected in their community can foster open discussions to address misconceptions surrounding the legislative process. This increases public awareness, trust, and engagement through education and empowerment, including their interests and concerns in developing and implementing AI in healthcare. Potential barriers include a lack of funding to create targeted resources in different languages and logistics. In addition, the lack of volunteers to conduct the research, craft the outreach proposal, and execute the tasks can halt the outreach process.

Tackling this angle is where urgency can be created. If the magnitude is heavy enough, it can become a window of opportunity, merging the streams in problem framing, policy, and politics. Overall, these takeaways are set up to equip patients and health professionals with existing information and processes and only require creative organization for outreach.

Limitations and Future Research Recommendations

The research pool surrounding algorithmic biases in healthcare AI is limited and evolving compared to other healthcare topics due to the technology development and implementation surge over time. Furthermore, health regulations such as The Health Insurance Portability and Accountability Act (HIPAA), which ensure data privacy, data quality, and availability, are not fully accessible, leading to possible generalization of findings. If the timeline is extended for future research, involving relevant stakeholders as clients would be beneficial. The necessary information can be obtained through surveys and interviews to understand the organization's problem-framing and mitigation strategies. This fosters a collaborative space for future coalition buildings and networking opportunities.

Conclusion

The development and implementation of AI in healthcare are evolving without foundational and sustainable mitigating alternatives. The research aims to assess patients, healthcare professionals, AI developers/technology companies, and policy-making bodies' problem framings of algorithmic biases in healthcare AI. These varying framings are analyzed through the multiple streams framework, capturing the three streams, problem, policy, and political environments, to understand how policy change occurs. The analysis of the policy dynamics equips stakeholders with strategies for advocacy, partnerships, and community-oriented championships.

Overall, there is an overarching theme and differing angling of the problem framing. Patients underscore the underrepresentation in training data and the need for clinician oversight. Healthcare professionals' framing exhibits a lack of knowledge of using AI tools and the broader social inequities in the datasets. AI developers/technology companies attribute this to a lack of representation in the training data, human-centered design, and diversity in the AI developer workforce. The policy-making bodies focus on the lack of a comprehensive regulator as the issue is complex, from the problem formation to the full implementation, and there are existing statistical biases in training data. The problem is multifaceted but with an overarching alignment on the lack of diverse datasets perpetuating inequalities. Addressing historical existing biases embedded in the datasets is a wicked problem on its own, thus bringing the stakeholders back to the starting point with the differing problem framings delaying the mitigation plan. Overall, there is a lack of guiding frameworks to inform and navigate the complexities surrounding the technologies implemented in healthcare settings.

However, based on the analysis, there are key takeaways, particularly addressing the patients' and healthcare professionals' involvement in advocacy. They focus on a guiding framework for research to understand policy dynamics, opportunities to participate in initiatives at the local and state levels, and outreach strategies to build a coalition. All fostering innovative evidence-based advocacy to mitigate algorithmic biases in healthcare AI. This holistic approach educates and empowers individuals to be creative to champion community empowerment and drive systematic change.

Acknowledgments

I want to thank God, my family, and my friends for their unconditional support throughout the program. They have been a pillar of inspiration and empowerment. I thank the MPPA cohort for their insightful perspectives on health, technology, education, and government reforms. I thank the faculty in Sacramento State's Master's in Public Policy and Administration program for creating an innovative higher education experience. They have inspired the structure and content of my research, particularly Amal Kumar. I want to especially thank Ahrum Chang, my advisor, for providing timely guidance, constructive feedback, and reassurance throughout my research.

References

Abràmoff, M. D., Tarver, M. E., Loyo-Berrios, N., Trujillo, S., Char, D., Obermeyer, Z., Eydelman, M. B., & Maisel, W. H. (2023). Considerations for addressing bias in artificial intelligence for health equity. *NPJ Digital Medicine*, 6(1), 170–170. https://doi.org/10.1038/s41746-023-00913-9

Aquino, Yves Saint James. (2023). Making decisions: Bias in artificial intelligence and data-driven diagnostic tools. *Australian Journal of General Practice*, 52(7), 439–442. https://doi.org/10.31128/AJGP-12-22-6630

Bauer-Kahan, Rebecca. (2024). *Assemblymember Bauer-Kahan Introduces a Bill to Eliminate Bias in AI Decision-Making*. California State Assembly. Retrieved from https://a16.asmdc.org/press-releases/20240215-assemblymember-bauer-kahan-introduces-bill-eliminate-bias-ai-decision

Becker, Josh. (2023). *SB 721 California Interagency AI Working Group.* California State Legislature. Retrieved from https://leginfo.legislature.ca.gov/faces/billTextClient.xhtml?bill_id=202320240SB721

Bonta, Rob. (2022). *Attorney General Bonta Launches an Inquiry into Racial and Ethnic Bias in Healthcare Algorithms*. Office of the Attorney General, State of California. Retrieved from https://oag.ca.gov/news/press-releases/attorney-general-bonta-launches-inquiry-rac-

California Department of Public Health. (2015). *Introducing the Let's Get Healthy California "Innovation Challenge" Improving Community Health and Promoting Health Equity*. California Department of Public Health. Retrieved from https://www.cdph.ca.gov/Programs/OPA/Pages/NR15-058.aspx

Chen, Y., Clayton, E. W., Novak, L. L., Anders, S., & Malin, B. (2023). Human-centered design to address biases in artificial intelligence. *Journal of Medical Internet Research*, 25(1), e43251–e43251. https://doi.org/10.2196/43251

Choudhury, A., & Asan, O. (2020). Human factors: Bridging artificial intelligence and patient safety. *Proceedings of the International Symposium of Human Factors and Ergonomics in Healthcare*, 9(1), 211–215. https://doi.org/10.1177/2327857920091007

Copeland, B. J. (2019). *MYCIN Artificial Intelligence Program*. In Encyclopedia Britannica. https://www.britannica.com/technology/MYCIN

Dodd, Bill. (2023). *SCR 17 Artificial Intelligence*. California State Legislature. Retrieved from https://leginfo.legislature.ca.gov/faces/billTextClient.xhtml?bill_id=202320240SCR17

Hague, D. C. (2019). Benefits, pitfalls, and potential bias in health care AI. *North Carolina Medical Journal (Durham, N.C.)*, 80(4), 219–223. https://doi.org/10.18043/ncm.80.4.219

Head, B. W. (2022). *Wicked Problems in Public Policy*. SpringerLink. https://link.springer.com/book/10.1007/978-3-030-94580-0

ial-and-ethnic-bias-healthcare

Jiang, F., Jiang, Y., Zhi, H., Dong, Y., Li, H., Ma, S., Wang, Y., Dong, Q., Shen, H., & Wang, Y. (2017). Artificial intelligence in healthcare: past, present, and future. *Stroke and Vascular Neurology*, 2(4), 230–243. https://doi.org/10.1136/svn-2017-000101

Lamanna, C., & Byrne, L. (2018). Should artificial intelligence augment medical decision making? The case for an autonomy algorithm. *AMA Journal of Ethic*s, 20(9), E902-910. https://doi.org/10.1001/amajethics.2018.902

McCall, C. J., DeCaprio, D., & Gartner, J. (2022). The measurement and mitigation of algorithmic bias and unfairness in healthcare AI models developed for the CMS AI health Outcomes challenge. *medRxxiv*. https://doi.org/10.1101/2022.09.29.22280537

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–e0000278. https://doi.org/10.1371/journal.pdig.0000278

O'Connor, S., & Booth, R. G. (2022). Algorithmic bias in health care: Opportunities for nurses to improve equality in the age of artificial intelligence. *Nursing Outlook*, 70(6), 780–782. https://doi.org/10.1016/j.outlook.2022.09.003

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

Office of the National Coordinator for Health Information Technology. (2024). *Health Data, Technology, and Interoperability: Certification Program Updates, Algorithm Transparency, and Information Sharing (HTI-1) Final Rule.* U.S. Department of Health & Human Services. Retrieved from https://www.healthit.gov/topic/laws-regulation-and-policy/health-data-technology-and-interoperability-certification-program

Panch, T., Mattie, H., & Atun, R. (2019). Artificial intelligence and algorithmic bias: Implications for health systems. *Journal of Global Health*, 9(2), 010318–010318. https://doi.org/10.7189/jogh.09.020318

Rajkumar, A., Oren, E., Chen, K., Dai, A. M., Hajaj, N., Hardt, M., Liu, P. J., Liu, X., Marcus, J., Sun, M., Sundberg, P., Yee, H., Zhang, K., Zhang, Y., Flores, G., Duggan, G. E.,

Irvine, J., Le, Q., Litsch, K., … Dean, J. (2018). Scalable and accurate deep learning with electronic health records. *NPJ Digital Medicine*, 1(1), 18–18. https://doi.org/10.1038/s41746-018-0029-1

Reddy, S. (2023). Navigating the AI revolution: The case for precise regulation in health care. *Journal of Medical Internet Researc*h, 25(2), e49989–e49989. https://doi.org/10.2196/49989

Richardson, J. P., Smith, C., Curtis, S., Watson, S., Zhu, X., Barry, B., & Sharp, R. R. (2021). Patient apprehensions about the use of artificial intelligence in healthcare. *NPJ Digital Medicine*, 4(1), 140–140. https://doi.org/10.1038/s41746-021-00509-1

Rittel Horst W. J., & Webber, M. M. (1973). Dilemmas in a general theory of planning. *Policy Sciences*, 4(2), 155–169. https://doi.org/10.1007/BF01405730

Rodger, Cathy McMorris. (2023). *Chair Rodgers Opening Remarks on the Potential for AI in the Healthcare System*. Energy and Commerce Committee Chair Rodgers. Retrieved from https://energycommerce.house.gov/posts/chair-rodgers-opening-remarks-on-the-potential-for-ai-in-the-health-care-system

Straus, D. (2010). *How to make collaboration work: Powerful ways to build consensus, solve problems, and make decisions.* (16th ed. edition). ReadHowYouWant.

Straw, I. (2020). The automation of bias in medical artificial intelligence (AI): Decoding the past to create a better future. *Artificial Intelligence in Medicine*, p. 110, 101965–101965. https://doi.org/10.1016/j.artmed.2020.101965

Thais, S., Shumway, H., & Saragih, A. I.(2023). Algorithmic bias: Looking beyond data bias to ensure algorithmic accountability and equity. *MIT Science Policy Review*. https://sciencepolicyreview.org/2023/08/mitspr-191618004007/

Upadhyay, R., Knoth, P., Pasi, G., & Viviani, M. (2023). Explainable online health information truthfulness in Consumer Health Search. *Frontiers in Artificial Intelligence*, p. 6, 1184851–1184851. https://doi.org/10.3389/frai.2023.1184851

Vorisek, C. N., Stellmach, C., Mayer, P. J., Klopfenstein, S. A. I., Bures, D. M., Diehl, A., Henningsen, M., Ritter, K., & Thun, S. (2023). Artificial intelligence bias in health care: Web-based survey. *Journal of Medical Internet Research*, 25(4), e41089–e41089. https://doi.org/10.2196/41089

Wen, D., Khan, S. M., Ji Xu, A., Ibrahim, H., Smith, L., Caballero, J., Zepeda, L., de Blas Perez, C., Denniston, A. K., Liu, X., & Matin, R. N. (2022). Characteristics of publicly available skin cancer image datasets: a systematic review. *The Lancet. Digital Health*, 4(1), e64–e74. https://doi.org/10.1016/S2589-7500(21)00252-1

Westcombe M. (2007). Dialogue mapping: Building shared understanding of wicked problems [Review of dialogue mapping: Building shared understanding of wicked problems]. *The Journal of the Operational Research Society*, 58(5), 696–697. Palgrave Macmillan Press.