Requirements for Developing an Ethical AI Policy Framework in Healthcare Research: A Case Study

By

Prakriti Sarkar

THESIS

Submitted in partial satisfaction of the requirements for the degree of

MASTER OF SCIENCE

in

Health Informatics

in the

OFFICE OF GRADUATE STUDIES

of the

UNIVERSITY OF CALIFORNIA

DAVIS

Approved:

_____

Nicholas R. Anderson, Chair

_____

Mark Fedyk

_____

Mark Carroll

Committee in Charge

2024

Table of Contents

List of Figures & Tables

Acknowledgments

I could not have accomplished this goal without enormous support from my family, the UCD MHI Graduate Studies faculty and staff, and my wonderful colleagues at UC Davis Health.

First and foremost, I would like to express my sincere gratitude and deep appreciation for my research advisor and thesis chair, Dr. Nick Anderson, for his support and guidance throughout the graduate program. He took a genuine interest in my research and created opportunities for me. It has been an honor to be a part of the MHI program and to learn from a diverse and brilliant group of professionals. I am forever grateful to Dr. Mark Fedyk for his support throughout my program and beyond as the co-chair and mentor. I am thankful to Mark Carroll for his professional mentorship and support. I have had the opportunity to work with numerous professionals within UC Davis Health. Every one of them has contributed to my subject knowledge and the success of this project. I would like to express my respect and gratitude to UC Davis Health CTSC for giving me funding as a GSR and all the research team members, especially Christy Navarro, for her mentorship and support.

Finally, I could not have done any of this without the support of my loving family, who gave me time, support, and encouragement to complete the MHI program.

Abstract

Background

The advent of Artificial Intelligence (AI) in healthcare marks a transformative era characterized by enhanced diagnostic tools, personalized treatments, and efficient patient care. However, the integration of AI into healthcare systems introduces complex ethical dilemmas, necessitating the development of robust frameworks to navigate these challenges effectively. This thesis explores the intricacies of establishing ethical maturity frameworks in biomedical research, aiming to bridge the gap between technological advancement and ethical considerations.

Objective

The primary objective is to develop a comprehensive understanding of the current ethical frameworks in health AI, identify existing gaps and limitations, and propose solutions through a scoping review of literature, gap analysis, and summarizing findings from the Research Data Ethics Maturity Model Project (README) workshop.

We used a two-step process. Initially, we conducted a scoping review of 94 papers on AI ethics in healthcare to assess the existing landscape, next, we pinpointed where these ideas are missing the mark, and identified the deficiencies within these frameworks. Finally, we held a workshop where people could come together to think of new ways to deal with these ethical issues. This whole process helped us get a full picture of what's going on with AI and ethics in healthcare right now.

Methods

We found that there's not enough research on how AI and ethics cross paths in healthcare, which means we need to look into this more. The literature review highlighted key areas requiring attention, such as mitigating AI biases, achieving consensus on AI governance, and developing comprehensible and actionable regulations. Our deep dive showed us key areas that need work, like making sure AI doesn't have biases, getting people to agree on how AI should be governed,

and making rules that everyone can understand and use. The README workshop we had was a great place for people to share ideas; participants developed six use cases related to health AI, focusing on practical applications and ethical considerations like making sure AI is fair, improving health safety, and helping hospitals run better.

Conclusion

AI has the potential to improve healthcare, but we must be careful about its ethical use. By creating strong guidelines for using AI responsibly, we can ensure that it's used in ways that are open, fair, and respect everyone's rights. This study adds to the conversation on how to use AI in healthcare responsibly and sets the stage for more research and guidelines in the future. Our findings underlines the necessity for continuous dialogue and collaboration among stakeholders to foster an ethically sound integration of AI in healthcare systems.

1. Introduction

1.1. Background

In the last five years, artificial intelligence (AI) has become a household name; from students to politicians, everybody is talking about AI. Some people are excited about automation and the rise of powerful tools, but others are afraid of losing their jobs to AI. The ethical concerns surrounding AI have become a topic of widespread discussion and debate among various stakeholders, from researchers to physicians to policy-makers. But the question remains: what happened in the last 5 years that led to such a significant shift in public awareness and concern about AI ethics?

In the last five years, several incidents involving AI systems have garnered widespread attention and contributed to a significant shift in public awareness and concern about AI ethics. These incidents include instances of biased decision-making by AI systems, privacy breaches resulting from the collection and misuse of personal data, and the dissemination of misinformation and fake news generated by AI. Not only such incidents but with the innovations of tools like ChatGPT and Gemini, the potential for AI to generate realistic and convincing content within seconds. This has raised concerns about the potential misuse of AI-generated content, including deepfakes and disinformation campaigns[1]. These incidents have highlighted the ethical risks and potential harm that can arise from the use of AI. As a result, the general public and institutions like schools, colleges, and research institutes have become more aware of the need for responsible AI development and use, leading to increased discussions about the ethical implications of AI. Moreover, media coverage of high-profile cases like the Cambridge Analytica Scandal[2], where personal data was used without consent for political campaigns, has further fueled public concerns and reinforced the importance of ethical considerations in AI.

It's true that the increasing integration of AI in various industries, including healthcare, has raised concerns about the potential displacement of human workers. While it's undeniable that certain jobs may become obsolete due to automation, especially blue-collar jobs. According to the Pew Research Center, significant numbers of blue and white-collar jobs will be affected by the year 2025[3]. The development and implementation of AI also create new employment opportunities. For instance, AI is driving the growth of new professions, such as data scientists, machine learning engineers, and AI ethicists. These roles require a deep understanding of AI and its ethical implications, which in turn highlights the need for continuous education and upskilling. Therefore, Ethical considerations in AI are not only crucial for addressing the potential risks and current impacts of AI on society but also for the future. The principles of ethical AI are and should be developed in the future to ensure that AI systems are designed and implemented in a manner that aligns with human values and societal well-being. That's why, in this paper, we will begin with the basics. Our goal is to comprehend the concept of ethics in AI, identify the gaps in policies and current literature, and explore potential ways to address these gaps in the future.

1.1.1. What is Artificial Intelligence (AI), and why is it important in healthcare?

The concept of artificial intelligence, first named by John McCarthy in 1956, traces its origins to earlier ideas like Vannevar Bush's knowledge amplification system and Alan Turing's thoughts on machine intelligence [4]. Artificial intelligence (AI) has steadily advanced, becoming a transformative force across various sectors, including healthcare. The foundational definition of AI as the science enabling machines to perform tasks requiring human intelligence remains pertinent, particularly in the medical field [5]. The rise of AI in healthcare has been rapid and impactful. These technologies range from diagnostic sensors to sophisticated machine learning (ML) algorithms aiding clinical decisions. However, the integration of AI into healthcare has not

been without controversy, notably in terms of ethical considerations—a concern highlighted by incidents such as the collaboration between Google DeepMind and the Royal Free London NHS Foundation Trust, which led to a significant debate over data privacy and ethical standards[6].

The collaboration between DeepMind and the Royal Free London NHS Foundation Trust raised several ethical challenges and concerns. One of the main issues was the transfer of identifiable patient records across the entire Trust without explicit consent, which raised questions about patient autonomy and privacy. Additionally, the collaboration was criticized for giving Google and DeepMind undue and anti-competitive leverage over the NHS, potentially hindering the adoption and growth of beneficial technology.

The array of AI in healthcare is diverse, with applications including, but not limited to, remote monitoring devices, automated imaging interpretation, and virtual health assistants. These technologies have been envisioned as both a utopian enhancement to healthcare delivery and a dystopian replacement of human care providers. Despite these polarized views, the potential for AI to revolutionize healthcare by providing personalized, efficient, and accessible care is widely acknowledged. Nonetheless, this optimism is tempered by the need for a robust ethical framework to address the inherent risks and challenges posed by AI technologies, especially in sensitive areas like healthcare, where the stakes involve human lives and the sanctity of personal data.

The European Commission's definition of AI as systems exhibiting intelligent behavior by analyzing and acting autonomously in their environment is particularly relevant in healthcare contexts[7]. The healthcare environment includes various settings such as hospitals, nursing homes, and private residences—all of which necessitate a careful and ethical handling of data and patient interaction. The case of Google DeepMind underscores the complexity of balancing

technological advancement with ethical considerations in healthcare, revealing gaps in traditional medical ethics frameworks when confronted with the novel challenges posed by AI[8].

In recent years, the ethical challenges posed by artificial intelligence (AI) have prompted a surge in the development of AI ethics guidelines, with over 84 frameworks proposed by a diverse array of stakeholders. These include not only technology companies and consultancies but also regulatory bodies and governmental agencies. This proliferation of guidelines reflects a growing recognition of the need to address the ethical implications of AI across various sectors. It signifies a collective effort to ensure that AI technologies are developed and utilized in a responsible and accountable manner. This diverse range of stakeholders demonstrates the widespread acknowledgment of the importance of ethical considerations in the advancement of AI, reflecting a global commitment to shaping the future of AI in a way that aligns with ethical principles and societal values[9]. These frameworks generally share common principles such as transparency, fairness, and accountability, but their interpretation and application can vary significantly. Despite the good intentions behind these frameworks, there is concern over the integrity and effectiveness of their implementation in the real world [10]. In recent years, the integration of artificial intelligence (AI) into healthcare systems has raised significant ethical concerns and sparked intense debates. As healthcare providers and policymakers increasingly turn to AI technologies to improve patient care, reduce costs, and enhance the efficiency of healthcare systems, it becomes imperative to address the ethical implications of this technological advancement. The ethical considerations surrounding AI in healthcare span various levels of abstraction, including individual, interpersonal, group, institutional, and societal dimensions. These considerations encompass issues related to the reliability and interpretability of AI-driven decisions, concerns regarding fairness and equitable outcomes, and challenges related to traceability. Furthermore, the

incorporation of traditional and non-traditional sources of health data into AI-driven decision-making processes necessitates careful protection and harmonization to ensure privacy, security, and ethical use. As the ethical landscape of AI in healthcare continues to evolve, it is essential for policymakers, regulators, and developers to proactively address these ethical challenges to build public trust and ensure the responsible and beneficial integration of AI technologies in healthcare delivery[11].

1.1.2. How do you define ethics specifically contextual to healthcare AI?

Ethical considerations in health AI involve upholding principles and guidelines that govern the responsible utilization of artificial intelligence systems within healthcare. These ethical standards ensure that AI systems prioritize patient safety, privacy, transparency, fairness, and accountability throughout their design, development, and deployment processes to address potential risks associated with bias in algorithms, data privacy breaches, lack of decision-making process transparency as well as addressing the impact on patient outcomes and trust. Upholding values such as beneficence, non-maleficence, autonomy, justice, and privacy is essential for ensuring that AI technologies benefit patients while respecting their rights and autonomy to promote equitable healthcare outcomes[12].

The evolving nature of AI poses additional ethical concerns, including the lack of well-defined laws or regulations to address legal and ethical issues within healthcare settings.

The absence of clear guidelines may further contribute to the widening of the digital divide framework guiding AI-based decision support systems. This gap is particularly concerning given the ethical challenges inherent in AI, such as biases in algorithm development due to inadequate or poor-quality training datasets, patient privacy protection, and building trust among patients and healthcare professionals. Currently, the healthcare sector lacks a universally recognized integration of AI in healthcare and faces numerous challenges, including those inherent to machine learning

science, logistical hurdles in implementation, and barriers to adoption that need to be addressed through appropriate clinical and sociocultural pathways changes. Developers of AI algorithms need to be acutely aware of potential risks, including dataset shift, accidental fitting of confounders, discriminatory biases, difficulties in generalizing to new populations, and unintended negative impacts on health outcomes. It is crucial to develop information systems that can detect and address unfairness effectively.

The need for a universally recognized framework guiding AI-based decision support systems in healthcare is becoming increasingly apparent. With the rapid expansion of AI in healthcare, there is a critical necessity to address the ethical challenges associated with its integration. A key concern in the realm of AI in healthcare is the potential biases in algorithm development due to inadequate or poor-quality training datasets. The repercussions of such biases have been demonstrated in real-world examples, such as the case of the algorithm developed by UnitedHealth Group[13], which exhibited a severe racial bias against Black patients. The manifestation of such biases as a Social Determinant of Health directly impacted the health outcomes of the patients. This underscores the urgent need for frameworks that prioritize transparency, fairness, and accountability in the development and deployment of AI-driven healthcare systems.

1.1.3. The good and the bad side of AI in healthcare

The use of AI is increasingly becoming popular in the field of healthcare; AI is not only popular among clinicians but also among those individuals who adopted the "4P model of medicine" (Predictive, Preventive, Personalized, Participative)[14] where the patient's or individuals active participation becomes the key in this model for example: wearing wearable devices and using smartphones to track one's health. One of the branches of health AI that has transformed healthcare is Precision Medicine; this approach has shifted the medical paradigm from a one-size-fits-all strategy to one that is personalized, considering unique genetic, environmental, and lifestyle

factors. Precision medicine has greatly helped in the early detection of disease, tracking the progression of diseases, and providing personalized treatments to patients[15]. In recent years, Machine learning, a branch of AI, has gained a lot of popularity in healthcare. It is used to analyze large amounts of data and model building; machine learning is also used for recognizing patterns in data. Deep learning, a branch of Machine learning, is gaining more and more popularity in detecting complex diseases like Tumors; there is research that claims that deep learning can perform like humans or even better sometimes in diagnosing diseases.[16].

Now, coming to the bad part of AI in healthcare, AI presents a significant amount of risks that we must manage. For example, publicly available AI tools, like ChatGPT & Gemini, can generate highly convincing text that may bypass traditional plagiarism detection tools, which poses a real threat to the integrity of medical literature as well as medical research. For instance, if researchers rely on AI to write scientific papers, there is a risk that the generated information may not only be unoriginal but also incorrect and misleading.[17]

Moreover, AI's capacity to produce images and data that are indistinguishable from the original ones can lead to the fabrication of research findings. This is particularly dangerous in a field where accurate data is critical for patient care, drug development, and medical procedures. The ethical implications are equally alarming. The use of AI in creating research outputs challenges the core principles of academic integrity. There are a few research papers that even listed ChatGPT as one of the co-authors [18].

As AI becomes more sophisticated and accessible, the medical community must enhance its vigilance and develop new tools to detect AI-generated fabrications. This is crucial to preserving the reliability of medical research and upholding the ethical standards that govern scientific inquiry.

1.1.4. Impact on Health Informatics

Health informatics involves the application of both data and technology in a healthcare setting, making it essential to consider the ethical dimensions that these AI tools can impose. As AI becomes more and more sophisticated and integrated into health systems, its ethical implications—such as patient privacy, data & research integrity will become pivotal subjects of analysis. This thesis explores these dimensions, thereby contributing directly to health informatics by aiming to improve the responsible use and governance of AI technologies. In the coming years, as AI evolves rapidly, so must the policies that govern its use in healthcare settings, including patient treatment and medical research. As Health informatics professionals, we will often find ourselves at the forefront of advocating for and developing policies that manage the ethical use of these technologies. Therefore, by examining ethical frameworks and assessing AI's impact on healthcare practices, this research not only aligns with but also enriches the field of health informatics.

It is imperative to recognize that ethical principles alone may not be sufficient to mitigate the potential risks of AI-driven healthcare systems exacerbating disparities. To address these challenges, ethical considerations in health AI involve a holistic approach that includes[10]:

1. Developing and adhering to robust ethical frameworks and guidelines specific to AI in healthcare.
2. Ensuring transparency and explainability of AI algorithms and decision-making processes.
3. Implementing mechanisms to identify and mitigate algorithmic biases that could perpetuate unfair treatment or discrimination in healthcare decision-making. This includes ensuring diverse and representative datasets, regular monitoring for biases, and implementing fairness measures within AI algorithms.

4. Fostering collaboration between AI developers, healthcare professionals, policymakers, and patients to ensure that ethical considerations are integrated into the entire lifecycle of AI systems in healthcare, from design to deployment and beyond.

This collaborative approach can help address ethical concerns in AI, promote accountability, and ensure that AI technologies are aligned with the values and needs of patients and society at large. Overall, ethical considerations in AI healthcare are crucial for promoting equitable access to quality healthcare and addressing potential biases and disparities. As AI continues to advance and reshape the healthcare landscape, it becomes essential to navigate the ethical challenges that arise. Therefore, our main goal in conducting this scoping review was to thoroughly explore both the academic and grey literature in this emerging field. We aimed to gain a deeper understanding of the ongoing discussions and debates surrounding the ethical considerations of AI in healthcare. Additionally, we sought to pinpoint areas where the existing literature may have gaps.

2. Rationale of Study

In this paper, we address the significant unmet needs to identify and evaluate the advent of Artificial Intelligence (AI) in healthcare, marking a transformative era that promises both immense benefits and novel challenges. AI's prowess in processing and analyzing vast medical datasets heralds a future of improved diagnostic precision and personalized treatment regimens. However, the rapid assimilation of AI into critical healthcare processes surfaces significant ethical concerns, necessitating a framework that ensures these technologies are deployed responsibly and equitably.

With the healthcare sector on the brink of a technological revolution, AI's ability to streamline complex cognitive tasks in diagnosis and treatment beckons a paradigm shift from traditional healthcare modalities. Yet, this shift is shadowed by the intricacies of AI's "black box" nature, raising pivotal questions about transparency and trust in AI-driven decision-making[19].

Central to the ethical debate is the application of AI across various patient demographics. Research demonstrates that AI's algorithmic bias can perpetuate health disparities, thereby underscoring the need for a conscientious framework that emphasizes transparency, equity, and answerability. The varying degrees of AI involvement among healthcare professionals highlight a spectrum of literacy and experience that influences the dynamics of ethical AI integration into healthcare systems.

The ethical frameworks currently evaluated by healthcare practitioners show the community's readiness to address AI's ethical dimension, balancing the scales between AI innovation and moral responsibility. Notably, graduate students forming the majority of the README workshop participants underscore the significance of cultivating an early understanding of these ethical concerns within future healthcare professionals.

As this thesis unfolds, it navigates through the realm of ethical AI in healthcare, inquiring into the multifaceted challenges that accompany the integration of AI into this sensitive domain. It advocates for a harmonious balance between technological advancement and ethical governance, with an underlying commitment to uphold human dignity, autonomy, and justice in healthcare delivery.

This thesis sets out to dissect the intricacies of ethical maturity frameworks in health AI, exploring the alignment, or lack thereof, between technological progress and ethical imperatives. It embarks on an interdisciplinary journey, scrutinizing the confluence of machine learning, clinical care, and ethical policy-making. Despite a surge in AI ethics frameworks, skepticism surrounds their application, integrity, and real-world effectiveness[20]. Disparities in AI applications—such as biases against minority groups in algorithms—highlight an urgent need for ethical frameworks that emphasize transparency, fairness, and accountability[21].

A pivotal part of this research involves evaluating community-engaged workshops, like the README project, that concentrate on converging ethical principles into coherent frameworks. Interestingly, graduate students constituted the majority of workshop participants, indicative of the rising interest and involvement of emerging scholars in shaping the ethical contours of health AI.

This study also navigates the undercurrents of policy-making and the implementation of ethical guidelines. Through a meticulous literature review and workshop findings, it aims to chart the course for future ethical frameworks in healthcare AI, advocating for a harmonious balance between AI innovation and ethical assurance. In essence, this thesis contributes to the scholarly discourse on health AI ethics and paves the way for practical guidelines, promoting an ethically attuned deployment of AI in healthcare.

3. Objective

Our study objectives are as follows:

- Aim 1. Scoping literature review for use and generation of health-related and clinical data
    - To better understand the different ethical frameworks in health AI, we conduct a Scoping Review of current literature for the use and generation of health-related and clinical data.
    - To identify and address the gaps and limitations of current ethical guidance and frameworks.
- Aim 2. Summarize data collected from participants during the README workshop
    - , i.e., (Findings of a workshop at UC Davis Health to leverage consensus elements of the protoype/s developed in the workshop to design and apply a framework.
    - 3.1. Create data visualization(s) to communicate the results of the consensus design thinking workshop
    - 3.2. Proposal and approach to implement/manage expectations for future management/training in the same area.

4. Methodology

4.1. Aim 1: Literature Review

The increasing integration of artificial intelligence (AI) and machine learning in products and decision-making processes has prompted a shift in public concerns from the misuse of personal data to the potential for biased or detrimental outcomes. This shift has led to a growing consensus on the necessity of AI regulation to ensure consumer trust and mitigate risks associated with opaque algorithms[18], emphasizing the need for businesses to anticipate and prepare for the impending regulations by deepening their understanding of the stakes, including the impact of outcomes and the trade-offs involved. Furthermore, the authors highlight the challenges companies will face in adhering to stringent AI explainability requirements, particularly in regions such as the European Union and the United States, where individual rights are highly valued. The need for global operators to navigate varying expectations for explanations further underscores the complexity of AI regulation on a global scale. As the EU takes the lead in proposing an AI legal framework[5], it is evident that regulation is deemed essential for fostering trustworthy AI tools.

The study by Obermeyer et al. (2019)[17] sheds light on the pervasive racial bias present in a widely utilized algorithm for healthcare management. The authors demonstrate that the algorithm systematically underestimates the healthcare needs of Black patients while overestimating those of White patients, leading to significant disparities in patient care. This racial bias has substantial implications for the allocation of healthcare resources and the overall well-being of affected populations. The findings of this research are consistent with previous studies that have highlighted the presence of bias in algorithmic decision-making processes. For example, few studies [22,23]have identified biases in algorithms, emphasizing the need for greater scrutiny of these systems to ensure fairness and equity. For example, machine learning applied to language corpora can create biases by learning and reproducing the cultural stereotypes present in the

data[22].The authors utilized word embeddings, which represent words as vectors in a high-dimensional space based on their contextual usage in text data. These embeddings capture the statistical regularities of language, including the associations between words and concepts. However, these associations can reflect and perpetuate societal biases, such as gender and racial stereotypes. Similarly, studies also address biases in facial analysis datasets and commercial gender classification systems[23]. It builds on prior research, such as the work by researchers on societal gender biases in Word2Vec[24]. The researchers demonstrate that existing datasets like IJB-A and Adience are overwhelmingly composed of lighter-skinned subjects, leading to significant disparities in the accuracy of classifying different gender and skin type groups. This bias is further perpetuated through machine learning algorithms, resulting in algorithmic discrimination.

In response to these concerning findings, there are a few researchers that have proposed several recommendations for addressing racial bias in healthcare algorithms[17]. Drawing on the work of De-Arteaga et al. (2019)[21], the author advocates for increased transparency and accountability in algorithmic decision-making processes, as well as the development of alternative models that prioritize fairness and accuracy. Similarly, there are studies that investigate the impact of demographic biases on face recognition algorithms. Research has shown that females and certain racial groups are more challenging to recognize[23]. Furthermore, the 2006 NIST Face Recognition Vendor Test revealed that algorithms developed in different hemispheres performed better on subjects of their respective regions. The study highlights the impact of training set distribution on algorithm performance, demonstrating how biases in training data can lead to decreased accuracy for certain demographic groups. These findings, as well as recommendations,

are crucial for mitigating the detrimental impact of biased algorithms on patient care as well as healthcare outcomes.

4.1.1. Research Question

Initially, the focus was on exploring literature related to concepts such as Artificial intelligence(AI), health, and ethics. However, as the review progressed, the focus shifted towards investigating the gaps in the literature and understanding the limitations in the existing literature, as mentioned by the authors. Articles were included if they addressed all four core concepts (AI, ethics, health, and biomedical) and were written in English. Enhancing the review's quality. Articles solely focusing on big data without explicit mention of AI methods were excluded, as were non-peer-reviewed academic materials,books, and irrelevant records like duplicates and incomplete entries

4.1.2. Criteria for Potentially Including Studies in this Review

In the rapidly evolving landscape of biomedical Artificial Intelligence (AI), where groundbreaking technologies are reshaping the future of healthcare, ethical considerations are more important than anything. As AI applications are used in various facets of healthcare, from diagnostics and treatment to personalized medicine, the need for robust ethical frameworks becomes increasingly pressing. Ethical decision-making in the realm of biomedical AI ensures the responsible development, deployment, and use of these technologies, safeguarding against potential biases[paper], ensuring equity, and upholding patient safety and trust.

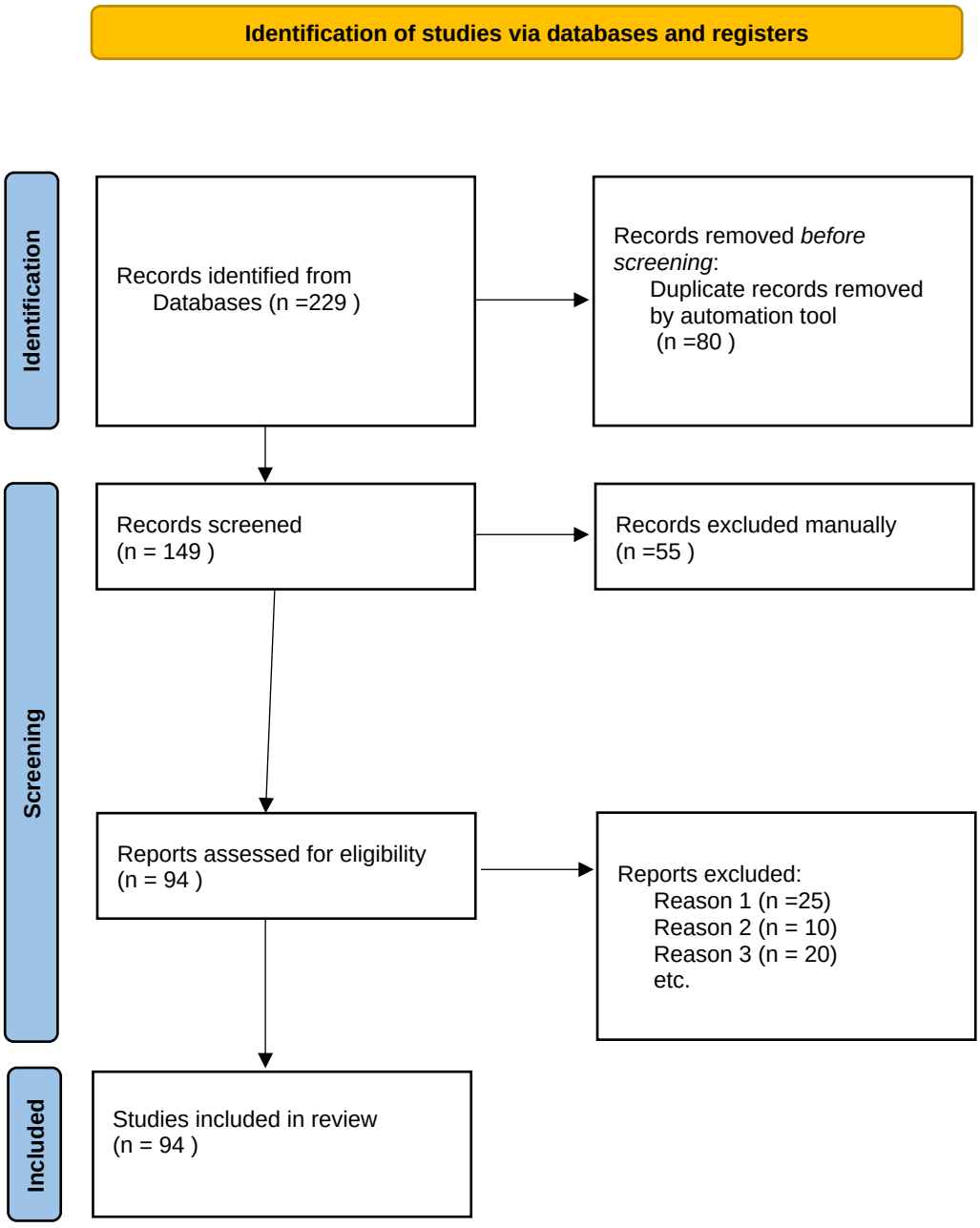
However, despite the critical importance of ethical guidelines, our exploration in reputable databases, such as Pubmed, IEEE, Scopus, and JSTOR databases, yielded a concerning gap in the literature. A search using specific keywords such as "biomedical," "ethics," "AI", and “Health” failed to yield any relevant studies. This absence shows a significant research void in the

intersection of these vital domains. Even when expanding the search parameters to include terms like "maturity model," "healthcare," "AI," and" ethics" combined together, the search remained unfruitful. This absence of studies in the last five years (2018-2023)—both in the broader context of biomedical ethics in AI and in the specific context of maturity models in healthcare AI ethics—shows the relevance of our research.

The absence of pertinent literature in established databases implies that a substantial gap exists in our understanding of ethical challenges in biomedical AI, particularly concerning maturity models in healthcare. The dearth of studies signifies a critical need for comprehensive research and exploration in this domain. The absence of existing literature is not merely a knowledge gap but also a call to action. It signals that there is a significant opportunity to contribute valuable insights that can guide the ethical development and implementation of AI technologies in biomedical contexts.

By delving into the territory of biomedical AI ethics and maturity models in healthcare AI, we aim not only to identify existing ethical concerns but also to propose frameworks and guidelines that can fill this void.

For this scoping review, a method developed by Arksey and O'Malley (2005)[25] was used. The method involved multiple steps to understand each document in the review and formulate a comprehensive research question covering a wide range of literature. The process followed a five-step approach outlined by Arksey and O'Malley (2005), complemented by suggestions from other researchers. The five-step approach of a scoping review is outlined in the figure below[Fig1].



Reason 1 – Date of publication
Reason 2 – Book chapters
Reason 3 – Non-peer-reviewed journals, abstracts, etc

Fig 1 Screening process of literature

4.1.3. Approach

The research followed a systematic approach inspired by established methodologies. We conducted a scoping review, utilizing methods recommended by Arksey and O'Malley's framework for scoping reviews. The core search concepts included AI, health, and ethics. Both

academic research articles and non-academic resources (grey literature) like white papers, govt articles& frameworks, policy papers, other agencies like Microsoft, CHAI, NIST, and even retracted papers were examined due to the dynamic nature of the AI field. Rigorous search techniques were applied to the grey literature to ensure a comprehensive review. The inclusion and exclusion criteria were determined and approved by the panel members. Articles were included if they addressed all four core concepts (AI, ethics, health, and biomedical) and were written in English. Enhancing the review's quality. Articles solely focusing on big data without explicit mention of AI methods were excluded, as were non-peer-reviewed academic materials, books, and irrelevant records like duplicates and incomplete entries.

4.1.4. Selection Criteria

A systematic search strategy was developed in collaboration with an experienced librarian. We conducted exhaustive searches across multiple databases, including PubMed, IEEE, Cochrane Library, Google Scholar, and Scopus, Proquest. The search terms were meticulously selected, encompassing concepts related to implementation, AI, and healthcare, like "artificial intelligence"**[MeSH Terms] OR "artificial intelligence"[tiab] AND ("ethical"[tiab] OR "ethically"[tiab] OR "ethics"[MeSH Terms] OR "ethics"[tiab] OR "ethic"[tiab] OR "ethics"[MeSH Subheading]).** We used standardized subject headings and word truncation to ensure a comprehensive search strategy

4.1.4.1. Inclusion Criteria:

Articles were included if they discussed AI, ethics, health, and biomedical comprehensively and were written in English. To ensure the relevancy and currency of the literature under consideration, we focused on publications(type) from the last five years **(2018-2023)**. This deliberate time frame was chosen to capture the rapid and dynamic advancements within the field of artificial intelligence

(AI) and its applications in healthcare. The past half-decade has witnessed an unprecedented acceleration in AI technologies, especially with the advent of groundbreaking innovations such as ChatGPT in 2022. The introduction of ChatGPT marked a significant milestone in the evolution of AI, particularly in the context of health AI and ethics.

The inclusion of this specific timeframe allowed us to capture the nuanced changes and evolving ethical considerations that emerged alongside the rapid pace of technological advancement. By focusing on the period from 2018 onwards, our scoping review provides a detailed and insightful examination of the most contemporary literature, incorporating the latest developments and perspectives in health AI, including the ethical challenges and opportunities introduced by innovative AI models like ChatGPT.

4.1.4.2. Exclusion Criteria

Articles solely focusing on big data or machine learning techniques without explicit mention of AI methods were excluded from the review. Similarly, non-peer-reviewed academic materials,books, and irrelevant records such as duplicates and incomplete entries were meticulously filtered out. The exclusion of materials lacking any of the four core elements (AI, ethics, health,and biomedical aspects) ensured the precision of the study. Additionally, non-English articles were excluded to maintain a uniform language criterion.

4.1.5. Search Strategy

The following sources were included in the search intended to generate literature for the scoping review:

Scopus – 2018 to 2023

PubMed – 2018 to 2023

IEEE – 2018 to 2023

JSTOR – 2018 to 2023

ProQuest – 2018 to 2023

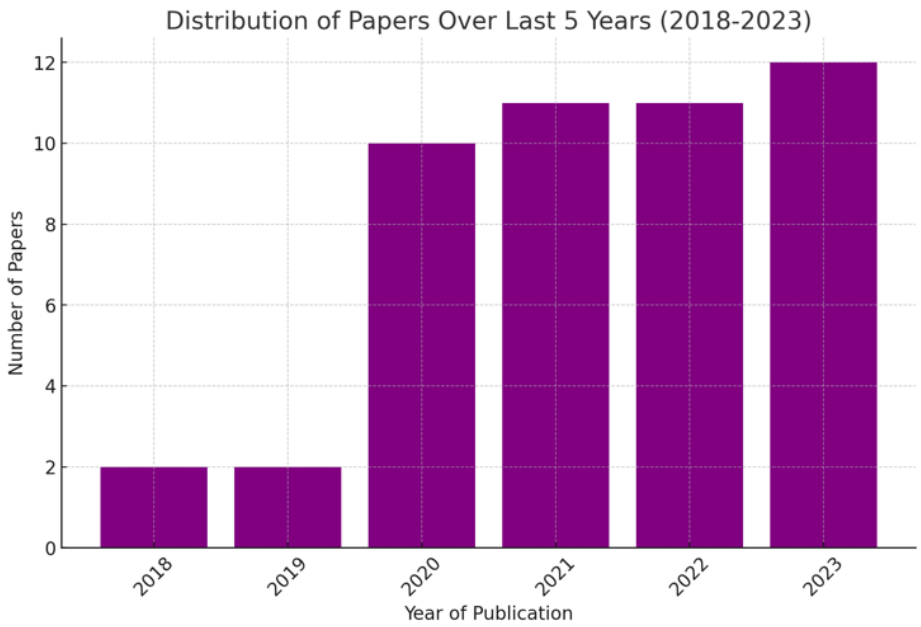


Fig 2: Distribution of papers according to the years of publication

4.1.6. Data Cleaning Process

Upon gathering the relevant literature, a systematic data cleaning process was initiated. The collected articles were reviewed, and relevant information was extracted. Keywords, themes, and clusters were identified as frameworks or evaluation of frameworks, allowing for a synthesis of the information gathered. The analysis focused on exploring the intersection of AI, health, and ethics, providing a holistic understanding of the current state of research in this multidisciplinary field. The clusters of papers are as follows:

Table 1: Clusters of retrieved papers

Cluster	Consists of
Forms (Types of Papers)	Opinion papers, case studies and case reports, conference papers.

Policy Papers	Government frameworks, regulatory papers, white papers.
Technology	Papers related to machine learning, deep learning, big data.
Related to Academia	Papers related to the publications.
Methodology	Papers related to scoping reviews, systematic reviews, mixed methods.

4.1.7. Documentation

A data extraction form was designed to be used to document the following from each study:

1. Study design
2. Method of analysis
3. Study objectives
4. Gaps found in research
5. The future direction of study
6. Study limitations
7. Definition and measurement of outcomes
8. Main findings

4.1.8. Research Participants

This literature review did not include direct research participants. The methodology included a scoping review of existing literature, each with its own sample pool.

No instrumentation was required for application among participants. Instead, a selection criterion was used to determine whether a study should be retained for inclusion in the systematic review. Tools like Zotero and Petal were used to track these studies.

4.1.9. Results

identifying and addressing the gaps and limitations of current research.

Key themes and gaps identified from the current literature:

Table 2: List of key gaps identified from current literature.

No.	Author and year of publication.	About the study.	Main gaps and themes identified.
1	Morley et al. (2022)[26]	Scoping review of literature and policy analysis conducted to identify underresearched policy areas related to AI-driven technologies in healthcare.	A significant gap in the international agreement on AI governance, which not only affects market competition but also poses challenges in developing effective policies that can benefit all stakeholders in healthcare.
2	Karimian et al. (2022).[27]	The authors conducted a systematic scoping review on the ethical issues of the application of artificial intelligence in healthcare.	The gaps that were highlighted in this study were the lack of practical methods for evaluating adherence to ethical principles in AI healthcare applications and the authors also mentioned the need for concrete guidance on preserving

			patient privacy during the developemnet of AI algoeithms.
3	NIST Face Recognition Vendor Test.[28]	Performance variations based on the origin of algorithm development.	Past reports from NIST have shown that most face recognition algorithms exhibit differential performance based on race, age, and gender. The FRVT focuses on technical performance and does not address the broader ethical and privacy implications related to the deployment of face recognition technologies
4	Obermeyer et al. (2019).[29]	The authors demonstrate that the algorithm systematically underestimates the healthcare needs of Black patients while overestimating those of White patients, leading to significant disparities in patient care. This racial bias has substantial implications for the allocation of healthcare resources and the overall well-being of affected populations.	sheds light on the racial bias present in a widely utilized algorithm for healthcare management.

5	Krijger, J. et al. (2023).[30]	This is an opinion paper that synthesizes literature and practical insights from mutual learning sessions with major organizations in the Netherlands to propose a holistic framework. This model outlines six crucial dimensions for operationalizing AI ethics in organizations, aiming to bridge the gap between theoretical AI ethics and practical implementation in governance and daily operations.	Identifies a significant gap in the literature concerning the operationalization of AI ethics within organizations. It highlights the lack of systematic approaches for advancing AI ethics procedures and integrating ethical principles effectively into organizational practices.
6	(Jobin, Ienca and Vayena, 2019).[31]	Scoping review of AI ethics guidelines, highlighting global convergence around five ethical principles and identifying divergence in	Despite the growing number of ethical guidelines, there's substantial divergence in how ethical principles are interpreted and implemented, highlighting the

		interpretation and implementation. Emphasizes inter-stakeholder cooperation, stakeholder engagement, and the importance of integrating guideline-development efforts with ethical analysis and adequate implementation strategies.	challenge of achieving a global consensus on ethical AI standards
7	Murphy et al. (2021).[32]	Ethical implications of AI in healthcare, interdisciplinary collaboration, privacy and security, bias and discrimination, accountability and transparency, informed consent and decision-making, social and cultural implications, and recommendations for future research in AI and healthcare.	Four ethical themes were identified as common across the health applications of AI addressed in the literature: data privacy and security, trust in AI, accountability and responsibility, and bias
8	Al-Hwsali, A. et al. (2023).[33]	Conducted a scoping review to explore the ethical and legal aspects of AI in Public Health.	The study shed light on important ethical and legal themes in AI for public health, like fairness, bias, privacy, and

		Reviewed 22 publications from 2015 to 2022, focusing on patient safety and privacy concerns.	accountability. It highlighted gaps in guidelines for responsible AI use, showing the need to address issues like health equity and privacy in AI technologies. The study stressed the importance of ensuring fair access to AI benefits, addressing privacy concerns from wearable devices, and the significance of having clear accountability frameworks for AI in healthcare.
9	European Commission. (2020).[34]	Need for a unified European approach to ensure trustworthy and ethical development of AI technologies, policy options for fostering a secure AI environment, addressing associated risks through regulatory and investment-oriented approach, and insights into the regulatory and ethical considerations essential for AI in healthcare.	Need for a unified European approach. Emphasis on addressing the associated risks of AI through a regulatory and investment-oriented approach

10	Ellefsen, A. P. T. *et al.* (2019).[35]	Discusses a cognitive gap in the literature related to Artificial Intelligence (AI) maturity models and Logistics 4.0, as well as the readiness of logistics companies to go digital and become smart and intelligent.	It also mentions a lack of literature dealing with the problem of AI maturity models and Logistics 4.0, indicating a gap in the existing literature.
11	(François Candelon et al., 2021).[36]	Authors emphasized the need for businesses to anticipate and prepare for the impending regulations by deepening their understanding of the stakes, including the impact of outcomes and the trade-offs involved . Furthermore, the authors highlight the challenges companies will face in adhering to stringent AI explainability requirements, particularly in regions such as the European Union and the	The costs and opportunities lost by companies in complying with AI regulations, and the need for companies to develop formal AI policies with commitments to safety, fairness, diversity, and privacy.

		United States, where individual rights are highly valued	
12	(Goirand et al., 2021).[37]	Conducted systematic scoping review methodology to investigate the implementation of ethics frameworks in AI-based Healthcare Applications (AIHA). The researchers conducted a comprehensive search of both peer-reviewed and grey literature related to ethics frameworks in AI applications in healthcare published between 2015 and 2020.	The study found that there are areas that need improvement when it comes to evaluating how well ethics frameworks are put into practice in AI-based healthcare. It also highlighted the importance of being proactive and involving everyone in the process to ensure ethical standards are met. Additionally, there is a challenge in making sure there is a fair balance of power between those providing AI healthcare services and those receiving them. These gaps show that there is still work to be done to make sure AI healthcare applications are ethically sound and beneficial for everyone involved

4.1.10. Summary

This table (Table 2) provides an overview of various studies addressing the ethical and policy-related implications of AI in healthcare, highlighting significant gaps and common themes across

the research. several common gaps are identified among these papers. One recurring gap found is the lack of a unified international agreement on AI governance, which creates challenges in developing policies that can be accepted and applied universally and benefit all stakeholders in healthcare. Another significant gap found is the presence of bias and discrimination in AI algorithms. Studies shows how AI can perform differently based on race, age, and gender, leading to inequitable healthcare outcomes. This indicates a critical need for addressing and mitigating biases to ensure fair treatment across all patient populations.There is a lack of practical methods for evaluating adherence to ethical principles in AI healthcare applications. This includes the need for concrete guidelines to preserve patient privacy during AI algorithm development. The divergence in interpreting and implementing ethical guidelines is also a common gap identified by the resaerchers. Addressing these gaps can lead to more ethical, fair, and effective AI applications in healthcare.

4.2. Aim 2: Evaluate community-engaged participant workshop for ethics principles

4.2.1. Background

As the README team of researchers, we aimed to develop an ethical response to the deployment of AI technology in healthcare settings, focusing on applications that would assist in the translation process. However, we realized that the overwhelming amount of literature produced on AI technology, its ethics, and its applications in the health sciences and the speed at which it was produced outmatched our small team of researchers and administrators. To address this challenge and to fulfill README and CTSC's (sponsor) commitment to training the next generation of translational data scientists, we formulated a strategy to build a community of ethicists who would trust each other and work together to manage the deployment and utilization of AI technology at our institution. We organized a workshop to facilitate the formation of this community, drawing

from a broad group of trainees, clinicians, researchers, and staff. Our goal was to create a team of ethicists who could cooperate with each other to develop something similar to a maturity model for the ethical deployment of AI technology.

4.2.2. Methods

The planning, designing, and successfully executing the workshop was divided into three phases those are 1) Pre-workshop, 2) During the workshop, and 3) Post-workshop.

4.2.2.1. Pre-workshop phase

Initially, we conducted a comprehensive review of existing ethical frameworks and pieces of literature on Health AI, ethical implications, and governmental policies and identified gaps in the literature. This step was useful in deciding the topics to discuss and present in the workshop to address these gaps.

Agenda: The workshop, themed around "Data Ethics for AI in Translational Science at UC Davis Health," was structured to foster deep understanding and collaborative development of ethical frameworks suited for AI applications in health research. Through the workshop, we aimed to bridge theoretical knowledge with practical application; the agenda of the workshop was to encourage participants to design, develop, test, and refine ethical models by brainstorming and collaborating among themselves in a short period of time.

4.2.2.2. Creating content:

The presentations and team activities were designed to address the specific ethical challenges identified in the field, providing participants with practical tools and frameworks to navigate these issues. The team developed the content with the help of the README project's aims, emphasizing actionable insights and strategies for enhancing ethical maturity. Additionally, a comprehensive

compilation of resources was prepared, offering participants a rich repository of materials to support their ongoing learning and application of ethical principles in their field of work.

4.2.2.3. Participant invitation

The organizing team held brainstorming sessions to identify and recruit participants. We created an Eventbrite link for smooth and free-of-cost registration. We then reached out to grad students from the Department of Health Informatics, Nursing School, and Medical School, we also invited researchers, faculties, IT professionals, and physicians.

We designed and launched a website(https://health.ucdavis.edu/ctsc/area/informatics/ethics-in-artificial-intelligence) the website functions as a comprehensive repository for ethical guidelines, research findings, and educational resources. Additionally, it highlights ongoing projects and collaborations, aiming to inspire community engagement and interdisciplinary dialogue. This platform also provides updates on policy developments and best practices, crucial for researchers navigating the complex landscape of AI ethics.

4.2.2.4. Stakeholder Engagement

Key stakeholders from various domains within the CTSC were engaged to provide the best workshop experience to the participants. We collaborated with different stakeholders to organize the venue and decided on the food menu to cater to all the dietary restrictions.

4.2.2.5. Workshop Execution method

The World Café method is an effective and adaptable format for facilitating large group dialogues, which is why we chose to utilize it in our workshop. This method is rooted in seven design principles that guide its implementation, ensuring that it can be tailored to meet various needs. It's been used in various settings, from community development to organizational change and even in strategic business workshops.

We integrated the World Café method into our workshop to foster meaningful conversations around the complex topic of ethical decision-making in health and AI. It allowed us to create a welcoming environment that encouraged participants to engage in intimate and constructive discussions. By breaking down the larger group into smaller groups of 5 to 6 participants, café-like table discussions, we provided a space where every voice could be heard, and insights could be built upon as participants moved between tables.

The choice of the World Café method was justified as it aligns with the collaborative spirit of our research aims. It is designed to draw out the "collective intelligence" of participants, ensuring diverse perspectives are heard and integrated.

4.2.2.6. Ethical Maturity Framework

The workshop introduced participants to the ethical maturity framework developed as part of the README research. This framework was a foundational element, guiding discussions and activities throughout the event. Some frameworks that we discussed were as follows:

4.2.2.7. Maturity Models

Maturity models[38] are tools organizations use to assess their current capabilities in specific areas, like technology adoption or project management. They help identify the organization's current stage of development, ranging from initial, where processes are unpredictable, to optimizing, where the focus is on continuous improvement. This framework guides organizations in evolving their practices systematically.

4.2.2.8. NIST Framework

Developed by the National Institute of Standards and Technology[39], this framework provides a structured approach to managing cybersecurity risks. It emphasizes the importance of creating

secure, resilient, and trustworthy AI systems, offering guidelines for ensuring that AI technologies are developed and managed responsibly.

4.2.2.9. Coalition for Health AI (CHAI)

This framework aims for a standardized approach to AI implementation to avoid inconsistent methods across the healthcare sector. It advocates for the creation of AI tools based on shared design principles that are ethical and secure and enhance user trust. The approach emphasizes broad collaboration, iterative guidance development, and the importance of considering a diverse array of stakeholder inputs.

CHAI[40] also focuses on critical checkpoints throughout the AI lifecycle to ensure the technology's performance remains aligned with changing demographics and scientific updates. Furthermore, it prioritizes the mitigation of algorithmic bias to uphold health equity, suggesting regular audits for fairness.

4.2.2.10. CARE Principles

These Care principles [41] advocate for Indigenous data sovereignty, emphasizing collective benefit, authority to control, responsibility, and ethics in data governance. They ensure that data about Indigenous communities is used in ways that respect their rights and interests, promoting ethical data practices.

4.2.2.11. FAIR Principles

Fair principles[42] aim to make scientific data more accessible and usable for the global research community. They ensure that data is Findable, Accessible, Interoperable, and Reusable, facilitating open science by making data easy to share and use across various research disciplines.

4.2.3. The Human Pangenome Project

We talked about the Human Pangenome Project[43] at our workshop because it's closely related to the second aim of our README research project. One of the primary goals of the README project is to ensure that the vast and complex data from such research is managed ethically and respectfully. We are creating a guide for scientists to handle this kind of information responsibly, ensuring that the rights and privacy of individuals are protected as we advance in health research. Therefore, we chose to spotlight the Human Pangenome Project at our workshop because it's closely related to README's second goal. The Human Genome Project marked a significant milestone in understanding the blueprint of human life. The discussions emphasized that live genomes are dynamic entities constantly evolving within our cells. Traditional data representations of genomes often collapse this complex, three-dimensional reality into a simplified, linear, and static model. This simplification can be misleading, as it fails to capture the full diversity inherent in human genetic material.

In the workshop, we delved into the ethical frameworks necessary to navigate the complexities of representing global genomic diversity. Questions of sampling and representation, such as the criteria for diversity, contributions to reference variation, and the ethical implications of legacy sample reuse and immortal cell line creation, were thoroughly discussed. Therefore, the Pangenome Project is a prime example of how modern science can improve health outcomes, but it also brings up big questions about privacy and ethics.

4.2.4. Our approach:

4.2.5. Interactive Sessions:

The "World Café" method was implemented by dividing participants into small groups. Each group developed a use case on a specific aspect of ethical AI in health research. These groups

engaged in deep discussions, guided by facilitators, to explore and challenge existing ethical frameworks. Each group developed their own use cases and respective frameworks.

There were total 6 use cases developed by each group of participants, and those are as follows:

Table 3: Use- Cases developed by the workshop participants

Groups	Use- Cases
Group 1	Administrative Risk to Reduce Administrative Error: Care Operations Quality Improvement/Organizational Readiness
Group 2	Determine Liver Scan Type is Screening or Diagnostic from Clinical Notes
Group 3	Diagnostic System to Improve Timeliness & Accuracy
Group 4	Risk of Late Discharge Extended Stay Optimize Operations
Group 5	Risk calculator for NICU discharge.
Group 6	Generating patient specific information (missing data)

Participants then rotated between different groups, bringing insights from new groups to their original discussions. This approach encouraged networking and the exploration of different perspectives, making it particularly effective for complex topics like ethical AI in health research. Through multiple rounds of discussion, participants move between groups, cross-pollinating ideas and gaining diverse insights from each other, enriching the collective understanding and developing innovative solutions.

Throughout the workshop, participants were asked to take PollEverywhere surveys in between presentations and brainstorming sessions. The information collected was on participant

engagement, previous knowledge, experience in AI, experience with patient care, and learning progress.

4.2.6. Post-Workshop Process

Survey: Participants' feedback was collected to assess the workshop's impact on understanding and applying the ethical maturity framework.

4.2.6.1. My Perspective and Learning:

Reflecting on the workshop's interactive format, the integration of presentations, group discussions, and hands-on activities significantly enhanced my understanding of ethical AI applications. The session that stood out the most was the "World Café," where we collaboratively explored real-world scenarios. Through focused brainstorming, we effectively combined our diverse expertise to pinpoint challenges and develop innovative solutions efficiently. This collaborative effort demonstrated the power of collective problem-solving in addressing complex ethical issues within a short timeframe. This was one of the key takeaways.

4.2.7. Results

In the workshop survey, we observed that nearly a quarter of the attendees (22%) are actively involved in patient care at UC Davis Health, while a significant majority (78%) do not engage directly with patients, pointing to a wide array of roles within the organization that don't require patient contact. Regarding data handling, only a few (7%) manage patient data on an individual level, and a slightly larger group (12%) deals with aggregated data. A more notable group (34%) handles patient data at both the individual and aggregate levels, and almost half (46%) do not work with patient data at all, which indicates diverse responsibilities related to data management among the workforce.

The survey also shed light on the extent of AI interaction among the participants, where a considerable number (48%) have used AI without being involved in its underlying design. Roughly one in four (24%) have experience both using and developing AI tools, showing a blend of practical application with technological innovation. A smaller segment (19%) is familiar with AI but has not used it, and an even smaller portion (10%) have developed AI technologies but do not use them regularly.

The survey showed a balanced interest in several key ethical frameworks for AI. More than a quarter of the participants (27% each) explored the STANDING Together Draft Recommendations for Health Data Documentation and the CARE Principles for Indigenous Data Governance, while 24% and 20% considered the CHAI Blueprint for AI and the AI Ethics Maturity Model, respectively. Only a few (2%) reviewed the NIST Risk Assessment Framework.

4.2.8. Discussion

Exploring this project required prespectives of both participant and organizer: Stepping into the participant's shoes was an important learning experience during the event. I listened, shared, and synthesized ideas as we developed use cases and navigated through brainstorming sessions. The cross-pollination of insights as we shifted between discussions enriched my understanding and challenged my preconceptions.

My dual role also offered a unique perspective on usability design. While I contributed to creating the content from an organizer's point of view, participating in discussions allowed me to appreciate the usability of our resources from the audience’s point of view. This dual insight was invaluable. It highlighted the importance of designing for me, ensuring that the resources we develop are not only informative but also intuitively accessible.

4.2.9. Next steps and Future improvements:

Looking ahead, there is a pressing need to expand the community of ethicists and to continue refining the ethical maturity model. Incorporating more diverse perspectives, particularly from underrepresented communities, will be crucial in ensuring that our ethical frameworks are inclusive.

In my opinion, while the workshop was a significant step forward, there is much work to be done. We need to develop more robust mechanisms for integrating ethical considerations into the AI development process and for ensuring that these considerations are reflected in the actual deployment of AI technologies. In our workshop, we had a majority representation of graduate students, future workshops like ours should aim to involve a wider array of stakeholders, including patients, to truly capture the ethical concerns associated with AI in healthcare.

4.2.10. Results and Summary

This thesis synthesizes insights from a review of academic publications and grey literature, shedding light on the field of ethical maturity frameworks in health research. The literature mostly highlights the necessity for a globally recognized frameworks to navigate the ethical landscape introduced by the integration of AI into healthcare systems. Key ethical themes that are found in the literature are mostly equity, data privacy, algorithmic transparency, fairness, and the safeguarding of human dignity. These issues are not standalone but are deeply interwoven. For example, the implementation of AI technologies in healthcare and in biomedicine is intimately connected to algorithmic transparency, ensuring unbiased and fair decision-making processes, which, in turn, is critical for protecting patient privacy and upholding human dignity.

The ethical considerations discussed in the papers that are extracted for drafting the thesis reflect broader concerns within the healthcare AI ethics landscape. For example, the review by Obermeyer et al. (2019) highlights the significant issue of racial bias in healthcare algorithms, pointing towards the essential need for algorithmic fairness and transparency. Similarly, the collaboration between Google DeepMind and the Royal Free London NHS Foundation Trust raised pivotal ethical concerns, emphasizing the need for rigorous data privacy protections and the establishment of trust in AI applications within healthcare. Despite these shared concerns among most of the papers, there remains a huge gap in the literature as well as in policymaking regarding the development and implementation of comprehensive ethical maturity frameworks that can address these ethical challenges in a holistic manner.

The thesis advocates for a dynamic and inclusive approach to developing ethical maturity frameworks in healthcare research. Such an approach requires active engagement from a wide array of stakeholders, including policymakers, healthcare professionals, patients, and IT specialists. The discussions and README workshops summarized in this study highlight the critical need for collaborative efforts to refine and implement ethical maturity frameworks that are adaptable and responsive to the evolving AI landscape in healthcare. README workshop was a small effort made by a team of researchers and dreamers to create awareness among the community as well as engage them in brainstorming ideas. There is a significant opportunity for leveraging these frameworks to enhance public health initiatives, particularly in low- and middle-income countries, by ensuring equitable access to AI-enhanced healthcare solutions. There has been a lot done in this budding field, but we have a long distance to cover to reach the destination of fair, inclusive, comprehensive ethical framework in health AI.

5. Conclusion

This study has revealed both the promise of AI and the ethical complexities it introduces. We have uncovered the need for multidimensional scrutiny where diverse insights lead to more robust and ethical AI applications.

Our journey into the AI landscape has highlighted the criticality of varied perspectives. A single reviewer's lens, while focused, is not enough to capture the kaleidoscope of ethical considerations that AI in healthcare demands. We've seen the importance of not just looking forward but also looking back, acknowledging the foundational works that have shaped the field. Through this scoping review, we have identified a significant gap in the existing literature on ethical challenges and maturity models in healthcare AI and emphasized the urgent need for comprehensive research to address this gap.

The intimate workshop setting provided a seedbed for rich discussion, yet we must plant these seeds in more diverse soils. Our findings are a clarion call to widen the circle of conversation, to bring more voices to the table – voices from different cultures, disciplines, and experiences – to ensure that AI is an ally in healthcare, accessible and beneficial to all.

This study is a stepping stone, not the final destination, in the pursuit of a future where AI and ethics walk hand in hand. The path ahead requires collaboration, continued dialogue, and an unwavering commitment to an inclusive vision. The ethical deployment of AI in healthcare is not an endpoint but a continuous process, one that we must approach with both caution and optimism. The ethical challenges posed by AI in healthcare demand comprehensive research and the development of frameworks and guidelines to address these complexities.

6. Study Limitations

The study has several limitations. The most significant is having only one reviewer evaluate the literature and primary data. Additionally, only articles published in the last five years were included.

The workshop faced several constraints that suggest areas for future improvement. First, 40 participants attended the workshop, which, while fostering a focused and intimate discussion environment, limited the scope of community engagement and awareness we could achieve. Expanding our participant base is important for a broader impact.

Secondly, we packed a significant amount of information into a limited timeframe, the restriction to use articles published within the last five years, while intended to capture the most recent discussions and incidents in AI ethics, inadvertently omits several past works and longstanding debates that might offer essential context to ongoing ethical considerations in AI development and application. Future sessions could benefit from a more streamlined agenda that allows for deeper exploration of fewer topics.

Lastly, the workshop participant's demographic was predominantly graduate students from UC Davis. While their insights were invaluable, a more diverse range of perspectives from professionals across different universities and organizations would enrich the exchange of ideas. Participation from various fields would better position us to drive meaningful change in the ethical use of AI in health research.

7. Future Direction

Our next steps are as critical as our findings. We must strive to refine our methodologies, broaden our workshops, and, most importantly, keep the conversation about ethical AI in healthcare dynamic, inclusive, and ongoing, involving more stakeholders from different institutions and from different fields. AI is involved rapidly, and so must our frameworks for understanding and guiding its use, ensuring that the healthcare of tomorrow is as compassionate as it is innovative.

8. References

1. Dhs.gov. Retrieved May 24, 2024, from https://www.dhs.gov/sites/default/files/publications/increasing_threats_of_deepfake_identities_0.pdf
2. Confessore, N. (2018, April 4). Cambridge Analytica and Facebook: The scandal and the fallout so far. *The New York Times*. https://www.nytimes.com/2018/04/04/us/politics/cambridge-analytica-scandal-fallout.html
3. Smith, A. (2014, August 6). *AI, robotics, and the future of jobs*. Pew Research Center. https://www.pewresearch.org/internet/2014/08/06/future-of-jobs/
4. McGugan, C. (2019) *What we thought AI was going to be, and what it has become, Forbes*. Available at: https://www.forbes.com/sites/forbestechcouncil/2019/06/26/what-we-thought-ai-was-going-to-be-and-what-it-has-become/?sh=ef4f10739ffe (Accessed: May 25, 2024)
5. Morley, J. et al. (2020) "The ethics of AI in health care: A mapping review," SSRN Electronic Journal. Doi: 10.2139/ssrn.3830408.
6. Powles, J. and Hodson, H. (2017) "Google DeepMind and Healthcare in an Age of Algorithms," Health and Technology, 7(4), pp. 351–367. doi: 10.1007/s12553-017-0179-1.
7. *Europa. eu*. Available at: https://www.europarl.europa.eu/RegData/etudes/STUD/2020/641547/EPRS_STU(2020)641547_EN.pdf (Accessed: December 26, 2023).

8. Jobin, A., Ienca, M. and Vayena, E. (2019) "The global landscape of AI ethics guidelines," Nature Machine Intelligence, 1(9), pp. 389–399. doi: 10.1038/s42256-019-0088-2.
9. Morley, J., Cowls, J., Taddeo, M., & Floridi, L. (2020). Ethical guidelines for COVID-19 tracing apps. Nature, 582(7810), 29–31. https://doi.org/10.1038/d41586-020-01578-0
10. Möllmann, N. R. J., Mirbabaie, M. and Stieglitz, S. (2021) "Is it alright to use artificial intelligence in digital health? A systematic literature review on ethical considerations," Health informatics journal, 27(4), p. 146045822110523. doi: 10.1177/14604582211052391
11. Washington Post (Washington, D.C.: 1974) (2019) "Racial bias in a medical algorithm favors white patients over sicker black patients," 24 October. Available at: https://www.washingtonpost.com/health/2019/10/24/racial-bias-medical-algorithm-favors-white-patients-over-sicker-black-patients/ (Accessed: December 27, 2023).
12. Grech, V., Cuschieri, S., & Eldawlatly, A. (2023). Artificial intelligence in medicine and research – the good, the bad, and the ugly. Saudi Journal of Anaesthesia, 17(3), 401. https://doi.org/10.4103/sja.sja_344_23
13. Akhtar, A. (2019) "New York is investigating UnitedHealth's use of a medical algorithm that steered black patients away from getting higher-quality care," *Business Insider*, 28 October. Available at: https://www.businessinsider.com/an-algorithm-treatment-to-white-patients-over-sicker-black-ones-2019-10 (Accessed: May 25, 2024)
14. Grech, V., Cuschieri, S., & Eldawlatly, A. (2023). Artificial intelligence in medicine and research – the good, the bad, and the ugly. *Saudi Journal of Anaesthesia*, *17*(3), 401. https://doi.org/10.4103/sja.sja_344_23

15. Candelon, F. et al. (2021) “AI Regulation Is Coming: How to prepare for the inevitable,” Harvard Business Review
16. Morley, J., Machado, C. C. V., Burr, C., Cowls, J., Joshi, I., Taddeo, M., & Floridi, L. (2020). The ethics of AI in health care: A mapping review. *Social Science & Medicine (1982)*, *260*(113172), 113172. https://doi.org/10.1016/j.socscimed.2020.113172
17. Obermeyer, Z., Powers, B., Vogeli, C., & Mullainathan, S. (2019). Dissecting racial bias in an algorithm used to manage the health of populations. *Science (New York, N.Y.)*, *366*(6464), 447–453. https://doi.org/10.1126/science.aax2342
18. Stokel-Walker, C. (2023). ChatGPT listed as author on research papers: many scientists disapprove. Nature, 613(7945), 620–621. https://doi.org/10.1038/d41586-023-00107-z
19. Savage, N. (2022). Breaking into the black box of artificial intelligence. Nature. https://doi.org/10.1038/d41586-022-00858-1
20. Grother, P., Ngan, M. and Hanaoka, K. (2019) Face recognition vendor test part 3:Demographic effects. Gaithersburg, MD: National Institute of Standards and Technology.
21. De-Arteaga, M., et al. (2019). Mitigating bias in algorithmic hiring: Evaluating claims and practices. arXiv:1901.09451.
22. Caliskan, A., Bryson, J. J., & Narayanan, A. (2017). Semantics derived automatically from language corpora contain human-like biases. Science (New York, N.Y.), 356(6334), 183–186. https://doi.org/10.1126/science.aal4230
23. Hardesty, L. (n.d.). Study finds gender and skin-type bias in commercial artificial-intelligence systems. MIT News | Massachusetts Institute of Technology. Retrieved May

24, 2024, from https://news.mit.edu/2018/study-finds-gender-skin-type-bias-artificial-intelligence-systems-0212

24. Buolamwini, J. (n.d.). *Gender shades: Intersectional accuracy disparities in commercial gender classification*. Mlr.Press. Retrieved May 25, 2024, from https://proceedings.mlr.press/v81/buolamwini18a/buolamwini18a.pdf
25. Arksey, H. and O'Malley, L. (2005) "Scoping studies: towards a methodological framework," International Journal of Social Research Methodology, 8(1), pp. 19–32. doi: 10.1080/1364557032000119616.
26. Morley, J., Machado, C. C. V., Burr, C., Cowls, J., Joshi, I., Taddeo, M., & Floridi, L. (2020). The ethics of AI in health care: A mapping review. Social Science & Medicine (1982), 260(113172), 113172. https://doi.org/10.1016/j.socscimed.2020.113172
27. Karimian, G., Petelos, E., & Evers, S. M. A. A. (2022). The ethical issues of the application of artificial intelligence in healthcare: a systematic scoping review. AI and Ethics, 2(4), 539–551. https://doi.org/10.1007/s43681-021-00131-7
28. Grother, P., Ngan, M., & Hanaoka, K. (2019). *Face Recognition Vendor Test (FRVT) part 2 :: Identification*. National Institute of Standards and Technology.
29. Obermeyer, Z., Powers, B., Vogeli, C., et al. (2019). Dissecting racial bias in an algorithm used to manage the health of populations. Science, 366(6464). DOI: 10.1126/science.aax2342
30. Krijger, J. et al. (2023) "The AI ethics maturity model: a holistic approach to advancing ethical data science in organizations," AI and ethics, 3(2), pp. 355–367. doi: 10.1007/s43681-022-002287.

31. Jobin, A., Ienca, M. and Vayena, E. (2019) "The global landscape of AI ethics guidelines," *Nature machine intelligence*, 1(9), pp. 389–399. doi: 10.1038/s42256-019-0088-2.
32. Murphy, K., Di Ruggiero, E., Upshur, R., Willison, D. J., Malhotra, N., Cai, J. C., Malhotra, N., Lui, V., & Gibson, J. (2021). Artificial intelligence for good health: a scoping review of the ethics literature. *BMC Medical Ethics*, *22*(1). https://doi.org/10.1186/s12910-021-00577-8
33. Al-Hwsali, A. et al. (2023) "Scoping review: Legal and ethical principles of artificial intelligence in public health," in Studies in Health Technology and Informatics. IOS Press.
34. *White Paper on Artificial Intelligence: a European approach to excellence and trust*. (n.d.). European Commission. Retrieved May 25, 2024, from https://commission.europa.eu/publications/white-paper-artificial-intelligence-european-approach-excellence-and-trust_en
35. Ellefsen, A. P. T. *et al.* (2019) *Logforum*, 15(3), pp. 363–376. doi: 10.17270/j.log.2019.354.
36. Candelon, F. (2021). AI Regulation Is Coming: How to prepare for the inevitable. Harvard Business Review.
37. Goirand, M., Austin, E., & Clay-Williams, R. (2021). Implementing ethics in healthcare AI-based applications: A scoping review. *Science and Engineering Ethics*, *27*(5). https://doi.org/10.1007/s11948-021-00336-3
38. Raza, M. (n.d.). *Maturity models for IT & technology*. Splunk. Retrieved May 25, 2024, from https://www.splunk.com/en_us/blog/learn/maturity-models.html

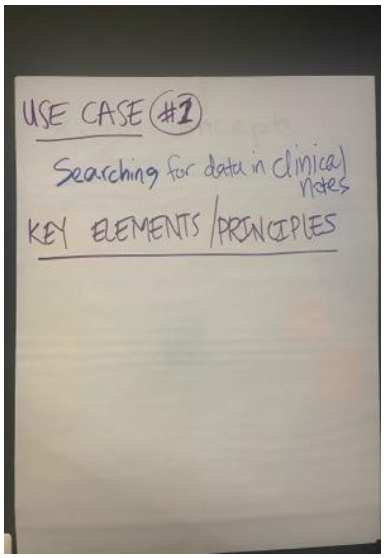
39. Nist.gov. Retrieved May 25, 2024, from https://www.nist.gov/itl/smallbusinesscyber/nist-cybersecurity-framework-0
40. CHAI. (n.d.). Coalitionforhealthai.org. Retrieved May 25, 2024, from https://www.coalitionforhealthai.org/
41. *CARE Principles* —. (n.d.). Global Indigenous Data Alliance. Retrieved May 25, 2024, from https://www.gida-global.org/care
42. Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L. B., Bourne, P. E., Bouwman, J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R., … Mons, B. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, *3*(1). https://doi.org/10.1038/sdata.2016.18
43. Wang, T., Antonacci-Fulton, L., Howe, K., Lawson, H. A., Lucas, J. K., Phillippy, A. M., Popejoy, A. B., Asri, M., Carson, C., Chaisson, M. J. P., Chang, X., Cook-Deegan, R., Felsenfeld, A. L., Fulton, R. S., Garrison, E. P., Garrison, N. A., Graves-Lindsay, T. A., Ji, H., Kenny, E. E., … the Human Pangenome Reference Consortium. (2022). The Human Pangenome Project: a global resource to map genomic diversity. *Nature*, *604*(7906), 437–446. https://doi.org/10.1038/s41586-022-04601-8

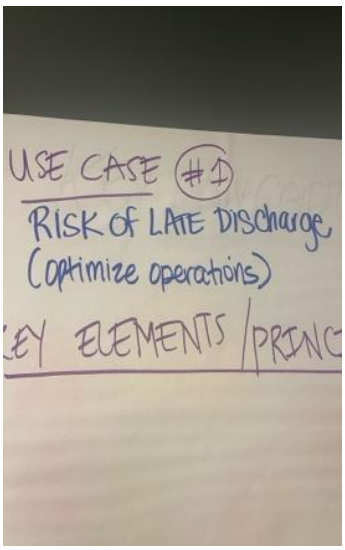
9. Appendix

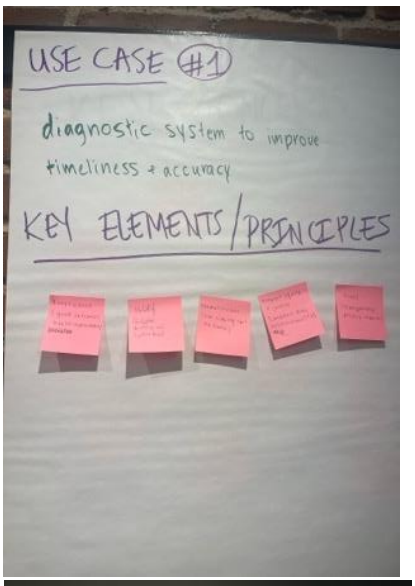
9.1. Questions asked to the README workshop participants using the POLL EVERYWHERE Survey are as follows:

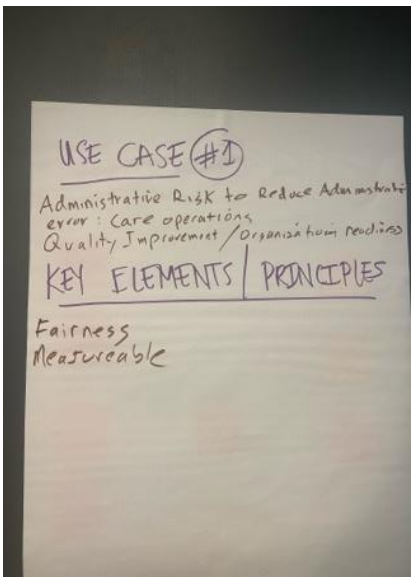
1. What is your role at UC Davis Health?
2. Does your work involve seeing patients at UC Davis Health?
3. Do you work with patient/clinical data?
4. To what extent have you worked with Artificial Intelligence?
5. In what context have you used AI at UC Davis Health?
6. What application(s) of AI at UC Davis Health have you been involved with, or are you aware of?
7. What were the key concepts you identified from this framework?
8. Which elements or concepts from this framework seem most relevant (to you) for an ethical framework for AI at UC Davis Health?
9. What elements, principles, or concepts do you believe are missing from those you identified in this framework?

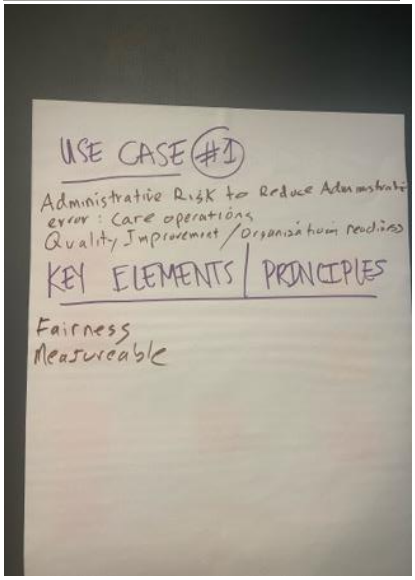
9.2. Pictures of Use Cases and frameworks designed by the workshop participants

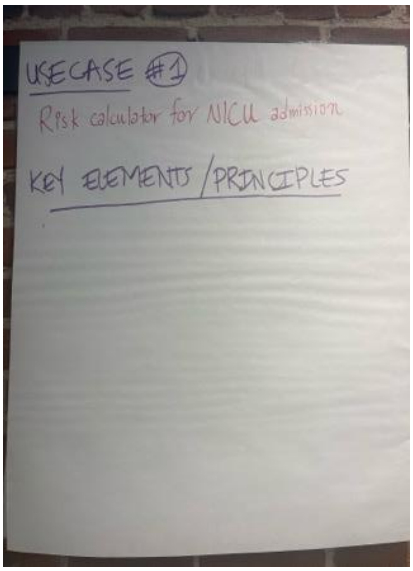


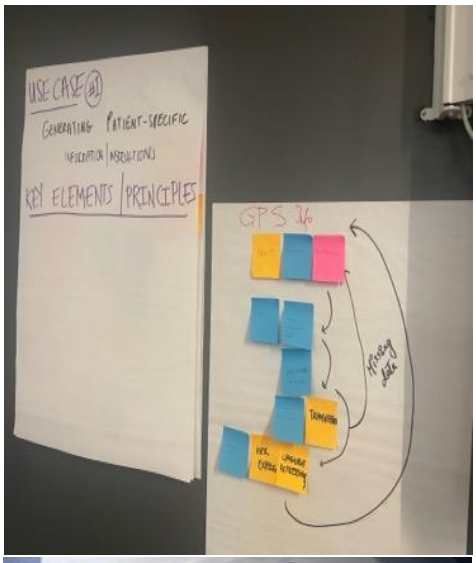


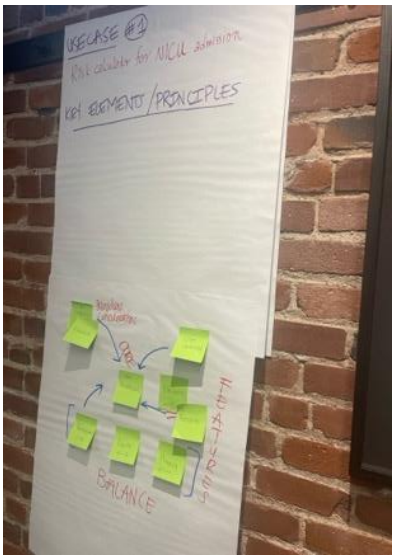

USE CASE #1
diagnostic system to improve timeliness + accuracy
KEY ELEMENTS/PRINCIPLES


USE CASE #1
KEY ELEMENTS | PRINCIPLES
Fairness
Measureable

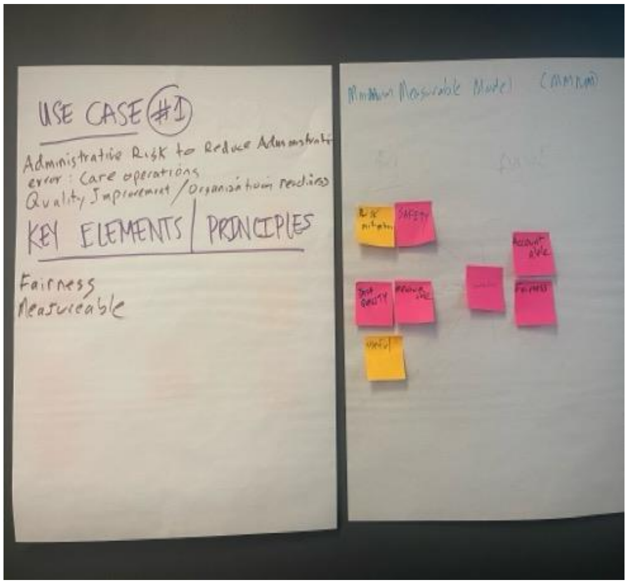

USE CASE #1
KEY ELEMENTS | PRINCIPLES
Fairness
Measureable

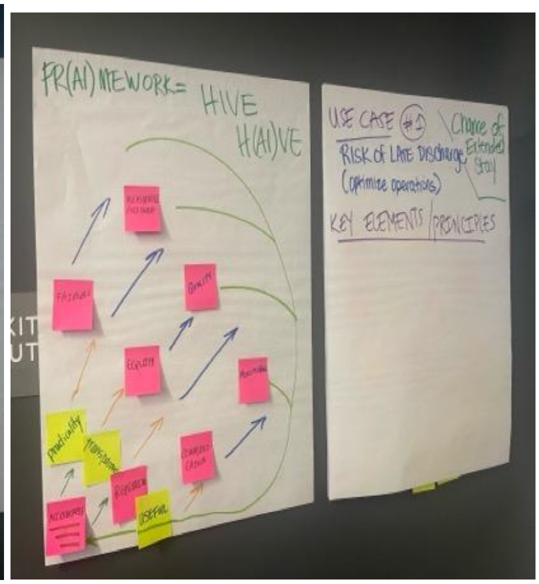

USECASE #1
Risk calculator for NICU admission
KEY ELEMENTS/PRINCIPLES


USE CASE #1
KEY ELEMENTS / PRINCIPLES
GPS
Training Data
Transition


USECASE #1
KEY ELEMENTS / PRINCIPLES
GPS
BALANCE


Determine Liver Scan Type is screening or diagnostic from clinical notes
KEY ELEMENTS / PRINCIPLES
Transparency
Repeat
NLP for Liver Scan Types


USE CASE #1
error: Care operations
KEY ELEMENTS / PRINCIPLES
Fairness
Measureable
SAFETY
FAIRNESS


FR(AI)MEWORK = HIVE
H(AI)VE
USE CASE #1
Chance of Extended Stay
RISK OF LATE DISCHARGE
(optimize operations)
KEY ELEMENTS / PRINCIPLES
practicality
USEFUL

9.3. Results: Workshop Findings and Data Visualization

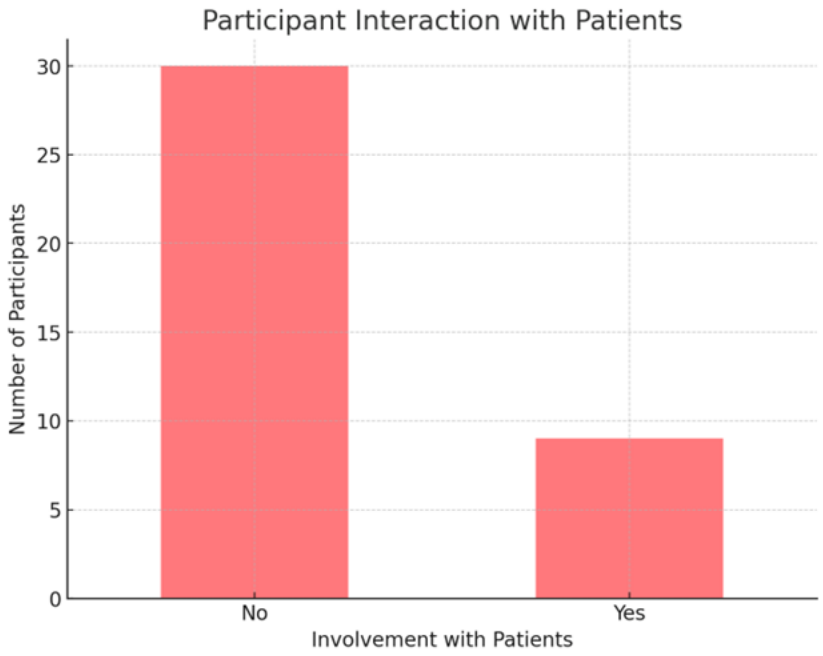


Fig 3 Participants involved in seeing patients at UC Davis Health versus those not.

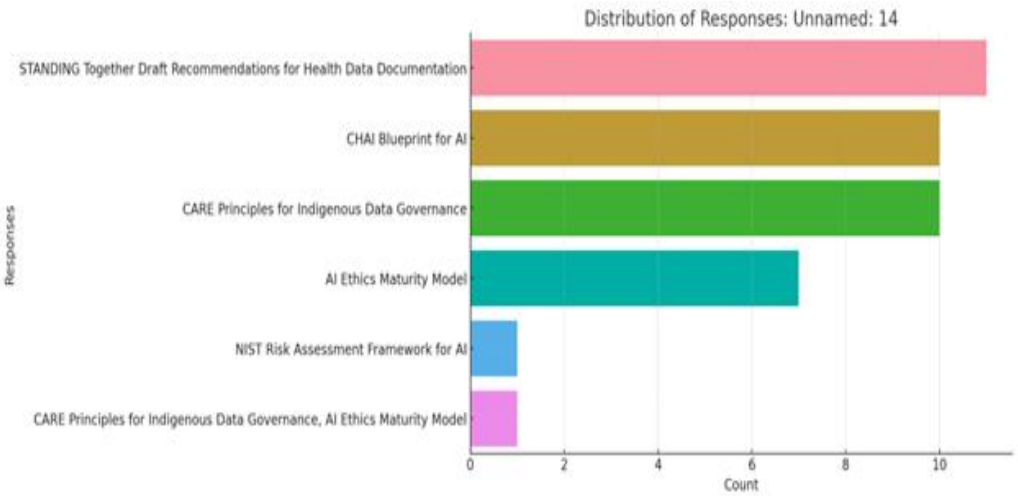


Fig 4: The different types of frameworks evaluated by the participants.

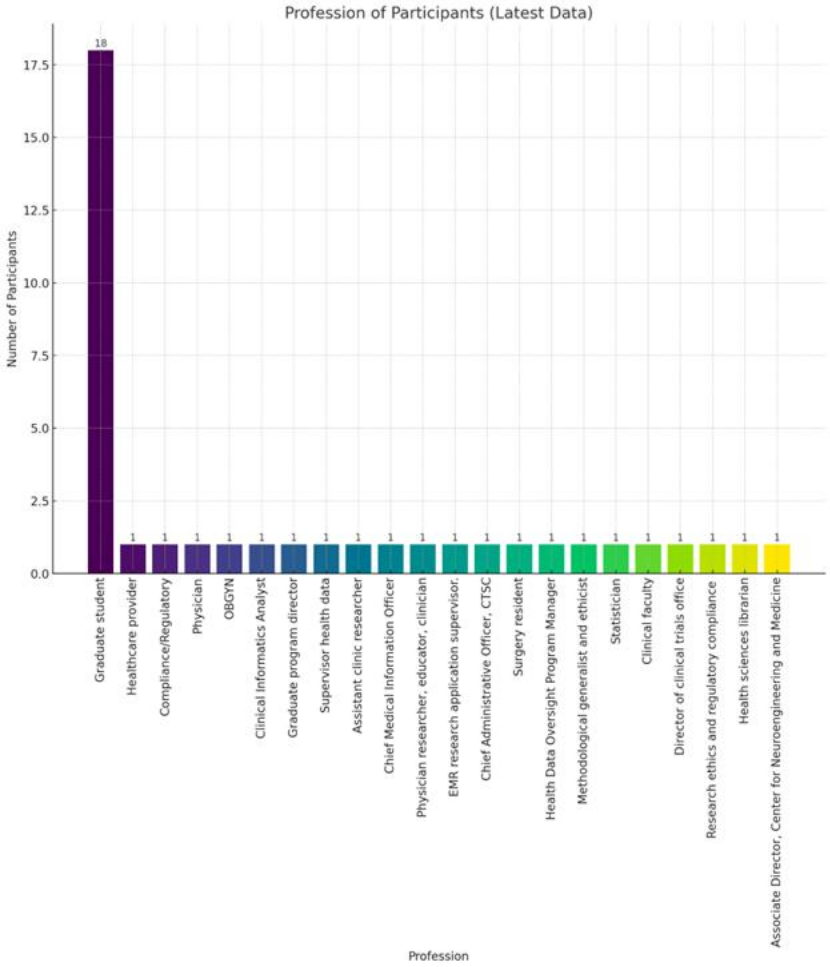


Fig 5: The professions of participants

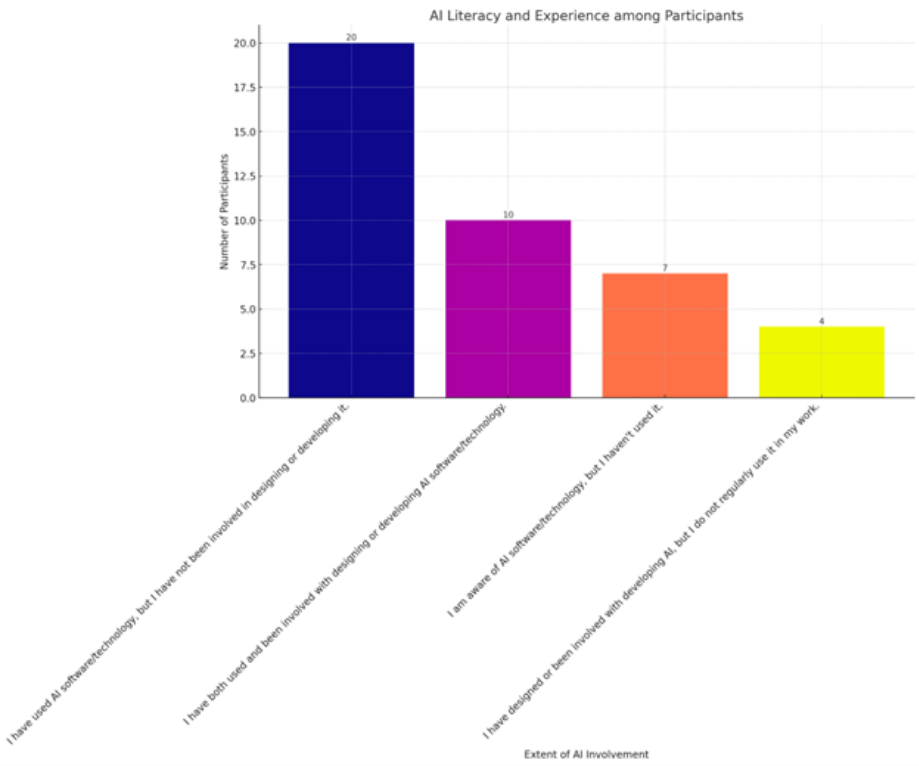


Fig 6: AI literacy and experience among participants

Fig 7: The contexts in which participants have used AI at UC Davis Health

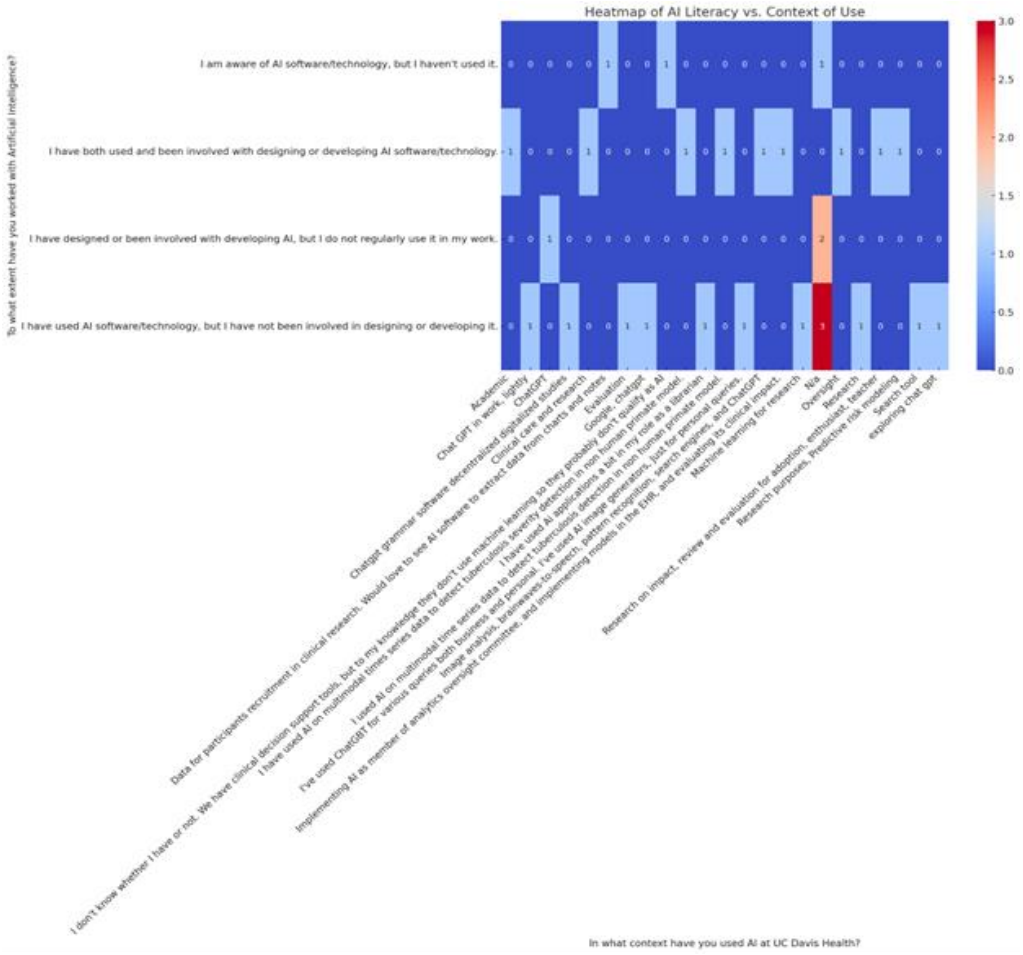


Fig 8: AI literacy and experience and the contexts in which they have used AI at UC Davis Health.

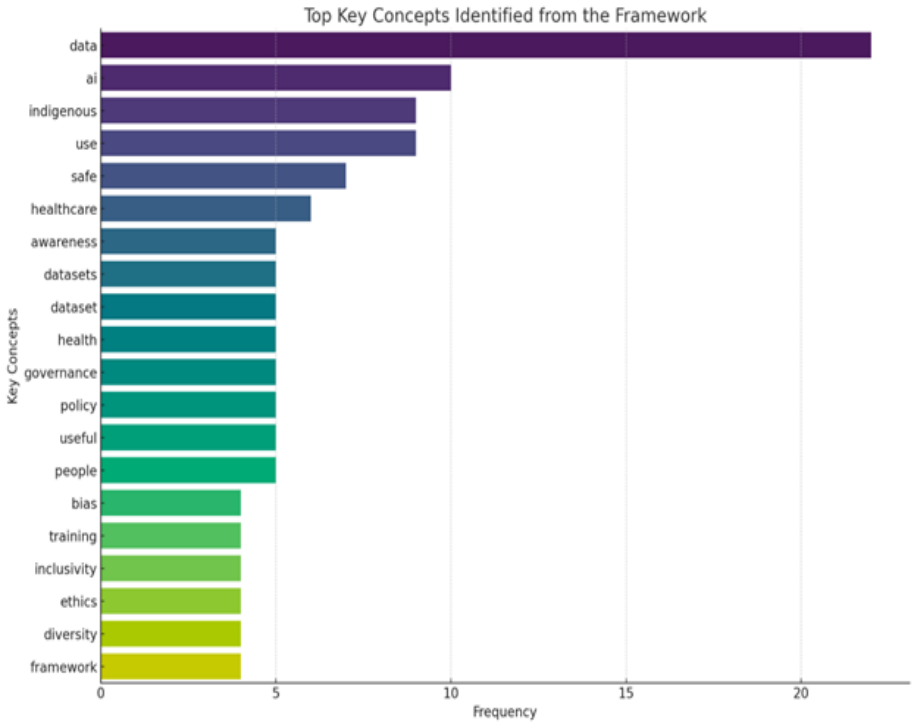


Fig 9: The top key concepts identified by participants from the framework based on the frequency of mentions in their responses