

Ethical Considerations in the Application of AI in Medical Research

Mohammed Lotfi^{1,2}

Department of Computer Science

¹*University of Amsterdam*

²*Vrije Universiteit Amsterdam*

Email: m.lotfi@student.vu.nl

Adam S.Z. Belloum^{1,2}

Department of Computer Science

¹*University of Amsterdam*

²*Netherlands eScience Center*

Email: a.s.z.belloum@uva.nl, a.belloum@esciencecenter.nl

Abstract—The use of Artificial Intelligence (AI) in healthcare offers significant improvements in diagnostics and patient care but also brings ethical challenges. This literature study examines key issues like privacy, data security, bias, transparency, and informed consent. It highlights the need to update ethical guidelines, engage various stakeholders, and work internationally to align AI practices. The paper also suggests conducting long-term studies to understand AI's impact on healthcare better. Addressing these points can support the medical community to responsibly use AI, ensuring it benefits patient care while maintaining high ethical standards. This review offers practical strategies for responsibly integrating AI into healthcare. This paper has been made in collaboration with researchers from the University of Amsterdam and the eScience Center, bringing together expertise from both healthcare and technology domains to enrich the study's findings and recommendations.

I. INTRODUCTION

The AI into medical research shows a new era of innovation, offering great opportunities to improve disease diagnosis, treatment personalization, and patient care. As AI technologies become increasingly widespread in this domain, they intend to revolutionize the way medical data is analysed and applied, potentially leading to significant breakthroughs in understanding and treating various health conditions [1] [2]. However, the rapid implementation of AI in medical research also brings forward a set of complex issues regarding ethical considerations that must be carefully navigated to ensure these technologies are deployed responsibly and for the greater good.

The ethical domain of AI in medical research is multifaceted, including issues related to privacy, data security, informed consent, algorithmic bias, and the potential for unintended consequences [3] [4]. Moreover, the dynamic nature of AI technologies, characterized by their ability to learn and adapt, poses unique challenges in ensuring transparency, accountability, and trustworthiness in their application [5]. Given the sensitive nature of medical research and the high importance of patient welfare, it is important to critically examine these ethical considerations to guide the development and implementation of AI systems in a manner that upholds the highest standards of ethical practice.

The term "considerations" in the title refers to the ethical issues that must be addressed when using AI in medical

research. Understanding these helps ensure AI is used responsibly and fairly, leading to better patient care and trust in medical technologies. This literature review aims to dive into the ethical dimensions of employing AI in medical research and applications, drawing upon a diverse set of articles that highlight both the potential benefits and ethical dilemmas associated with these technologies. By highlighting insights from these studies, the review seeks to provide a comprehensive overview of the current ethical debates surrounding AI in medical research, identify key themes and concerns, and offer recommendations for navigating the ethical challenges existing in this rapidly evolving field.

The goal of this paper is to help experts answer their research questions on the ethical considerations of AI in the healthcare domain and has been developed in close collaboration with field experts and researchers from the e-Science Center. The paper aims to provide a comprehensive understanding of the ethical challenges and offer practical strategies for responsible AI integration in medical research and patient care. This has added tremendous weight and depth to the ethical analysis of AI in medical research. Their collective insights guided this study to address both current and valid ethical aspects of the use of AI in medical research.

II. METHODOLOGY

To examine the ethical aspects of AI in medical research, an outline of research questions is made. These questions will outline the exploration of the key ethical issues, aid to understand the challenges posed by advanced AI algorithms, and look into the guidelines and frameworks that aim to address these concerns. This study is divided into two main research questions (RQ's) that form the basis and sub-research questions (SRQ's) that shed light onto different aspect of this study.

The development of these research questions was guided by an extensive literature review and expert collaboration. Collaborating with researchers from the e-Science Center and experts in AI and medical research, we identified key ethical issues and challenges. This approach ensured the questions are relevant, timely, and comprehensive, covering multiple aspects of AI ethics in medical research.

RQ1. What are the primary ethical concerns associated with the use of AI in medical research?

Understanding the primary ethical concerns is key to ensuring that AI technologies are developed and used in ways that complement societal values and healthcare standards. This entails issues like patient privacy, data security, consent, bias, and the ethical use of AI-generated knowledge [1] [4]. This question guides the ethical framework within which AI tools are created and operated, used to aim to protect human rights and ensure that AI aids rather than hinders or endangers medical advancements. To add onto RQ1, some sub research questions have been constructed.

SRQ1.1 What are the implications of data privacy and consent in the context of AI-driven data collection and analysis?

Data privacy and consent are essential in ethical medical research, especially as AI processes large amounts of data [10]. Ensuring the security and appropriate use of patient information is crucial to maintaining trust and compliance in healthcare. Addressing these concerns contributes to preserve the integrity of medical research and patient care [11].

SRQ1.2 How does the deployment of AI in medical research intersect with issues of bias and fairness?

Bias and fairness in AI systems are critical concerns in medical research and treatment, as they can lead to healthcare disparities [11]. Understanding how AI deals with these issues ensures that the technology contributes to fair healthcare. Addressing bias is crucial for developing AI that provides fair and consistent medical evaluations and treatments for all populations [12].

SRQ1.3 What ethical challenges arise from the potential lack of transparency and explainability in AI models used in medical research?

Transparency and explainability are essential in AI applications, especially in medical research where decisions significantly impact patient care. Understanding how AI models work is crucial for verifying their reliability, gaining clinician trust, and effectively using AI recommendations [13]. Additionally, explainability contributes in ethical audits and regulatory compliance, ensuring AI applications meet ethical standards and justify their decisions clearly [12].

RQ2. What frameworks or guidelines have been proposed or implemented to address these ethical concerns?

This question examines the frameworks and guidelines that address ethical concerns in AI within medical research. It is crucial to ensure AI adheres to standards that protect human rights and dignity, promoting transparency, accountability, and fairness. These guidelines are key to maintaining public trust and guiding the responsible use of AI technology [8].

SRQ2.1 What existing ethical frameworks and guidelines are being used, adapted or developed to address the challenges posed by AI in medical research?

Adapting and developing ethical frameworks for AI ensures responsible use, addressing bias, privacy, and moral issues in automated decisions. These guidelines help uphold ethical standards and protect patient welfare, fostering trust and integrity in medical research [1] [3].

SRQ2.2 How are regulatory bodies adapting policies and governance to the rapid integration of AI in medical research?

Regulatory bodies shape AI use in healthcare through policymaking and governance, ensuring AI applications are safe, effective, and equitable. Their actions guide how AI tools are developed and monitored, addressing risks and ethical concerns. This oversight tries to ensure that stakeholders meet legal standards and improves accountability in healthcare AI [8].

III. LITERATURE SEARCH AND SELECTION

The literature search consisted of a preliminary list of articles that served as a foundational reference point, called *Scopus Student*. This file has been constructed by expert researchers active in the field of AI, ethics and medical research, these researchers also participated in this study. They tested various queries to narrow the scope of the research and suggested keywords to limit the number of articles covered to answer the research questions. As a result, we selected only open-source articles in English from the last five years.

Using the LitStudy Python package [25], we performed data extraction on articles from the *Scopus Student* file. This process helped us refine the research direction and formulate research questions. The extracted data included insights into the articles' metadata, publication dates, most frequent authors, publication sources, and topic extraction/grouping, allowing us to analyze and focus our study effectively.

The selection of articles focused on those that dealt with the ethical issues of AI in healthcare and showcased the latest AI advancements and innovations. Each article was chosen for how deeply it explored ethical questions, how new and interesting its AI applications were, and how directly it related to medical practice and research. Keywords such as "artificial intelligence," "ethics," "medical research", were used as guidance in narrowing the scope. This led to the inclusion of studies like that by Hogg et al., which evaluated the safety and efficacy of AI in diagnosing medical conditions, thereby contributing to the ethical discourse around AI in patient care [5].

Other studies like, Weinert et al. explored stakeholder requirements for ethical AI implementation in medical settings, providing insights into the diverse perspectives that must be considered in ethical AI deployment [3]. Gundersen and Bærøe offered a forward-looking analysis on the future ethics of AI in healthcare, emphasizing the need for collaborative and deliberative approaches to ethical AI development and use

in medical contexts [4]. Also Notable among these initial references was the work of Al Kuwaiti et al. who provided a comprehensive review of AI's role in healthcare governance, highlighting critical ethical dimensions [1]. Complementing this perspective, Sallam (2023) delved into the use of AI models like ChatGPT in healthcare education and research, raising important ethical considerations [2].

In summary, the chosen articles present a unified view on the ethical issues related to AI in medical research. They stress the need for careful and principled use of AI in healthcare, underlining the importance of governance, engaging with stakeholders, and continuous ethical oversight as AI technology rapidly advances.

IV. RQ1. WHAT ARE THE PRIMARY ETHICAL CONCERNS ASSOCIATED WITH THE USE OF AI IN MEDICAL RESEARCH?

In the review paper by Al Kuwaiti et al, Figure 1 shows how AI is used in various healthcare areas [1]. These include medical imaging, diagnostics, virtual care, remote treatment, patient engagement, and support for health plans. AI also improves administrative efficiency, supports rehabilitation with advanced analytics, and enhances medical research and drug delivery by finding new patterns. Essentially, AI is revolutionizing healthcare for practitioners and patients alike.

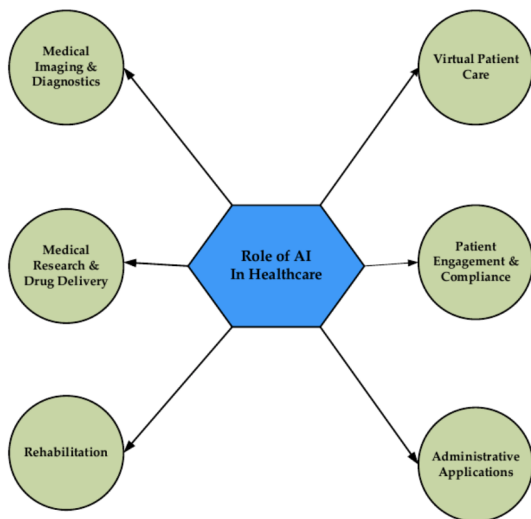


Fig. 1: Application of AI in various aspects of healthcare. [1]

Figure 2 presented by the WHO and shows an overview of the advantages and potential drawbacks associated with the use of language model mediators (LMMs) in various healthcare settings. It categorizes the use of LMMs into different sectors within healthcare, such as clinical care, patient engagement, administrative tasks, education, and research. For each category, the table contrasts the potential improvements that LMMs might bring against the risks and ethical concerns that need to be addressed to ensure their beneficial integration into health-related fields. [24]

Table 1. Potential benefits and risks in various uses of LMMs in health care

Use	Potential or proposed benefits	Potential risks
Diagnosis and clinical care	Assist in managing complex cases and review of routine diagnoses Reduce the communication workload of health-care providers ("keyboard liberation") Provide novel insights and reports from various unstructured forms of health data	Inaccurate, incomplete or false responses Poor quality training data Bias (of training data and responses) Automation bias Degradation of skills (of health-care professionals) Informed consent (of patients)
Patient-guided use	Generate information to improve understanding of a medical condition (as a patient or as a caregiver) Virtual health assistant Clinical trial enrolment	Inaccurate, incomplete or false statements Manipulation Privacy Less interaction between clinicians and patients Epistemic injustice Risk of delivery of care outside the health system
Clerical and administrative tasks	Assist with paperwork and documentation required for clinical care Assist in language translation Completion of electronic health records Draft clinical notes after a patient visit	Inaccuracies and errors Inconsistent responses depending on prompts
Medical and nursing education	Dynamic texts suited to each student's needs Simulated conversation to improve communication and to practise in diverse situations and with diverse patients Responses to questions accompanied by chain-of-thought reasoning	Contribute to automation bias Errors or false information undermine the quality of medical education New burden of learning digital skills
Scientific research and drug development	Generate insights from scientific data and research Generate text for use in scientific articles, manuscript submission or peer-review Analyse and summarize data for research Proofreading De novo drug design	Cannot hold algorithms accountable for content Algorithms encode bias towards the perspectives of high-income countries Generate information and/or references that do not exist Undermine key tenets of scientific research, such as peer review Exacerbate differential access to scientific knowledge

Fig. 2: WHO benefits / risks in uses of LLMs in health care [24]

In the literature study, insights from the articles in the student scopus file are used to examine the main ethical concerns linked to AI in medical research. These insights shed light on crucial ethical themes arising from AI's implementation in medical research, showcasing the set of ethical considerations brought about by integrating AI into this domain.

A. 4.1 Data Privacy and Security in AI Systems

In the paper of Al Kuwaiti et al. the critical issue of securing patient data within AI systems is highlighted. The ethical concern revolves around ensuring that patient data, often used to train AI models, is handled with high confidentiality and security to prevent unauthorized access or breaches [1].

In medical research and practice, patient data, ranging from medical histories and genetic information to real-time biometrics, is crucial for training AI models. These models can predict disease progression, recommend personalized treatment plans, and enhance diagnostic accuracy [9]. For example, an AI system might analyse thousands of patient records to learn patterns that predict the beginning of diabetes [9].

In healthcare, the integration of AI presents ethical challenges coming from the need for large data sets for machine learning, conflicting with patient confidentiality and the lack of willingness from healthcare organizations to share

sensitive data [10]. Security risks are increased as health records, valuable and vulnerable, often become targets for data breaches, highlighting the ethical importance to protect patient privacy. AI can also be over fitted which can lead to inaccurate medical predictions, and issues like data leakage further increase ethical concerns [14]. The non-transparent state of AI's decision-making processes, or the "black box" problem, challenges the transparency and trust necessary in healthcare.

The main risk in using patient data for AI training involves potential data breaches that could expose sensitive patient information, leading to privacy violations and potential misuse. Furthermore, if AI systems are not thoroughly tested and secured, they might be vulnerable to cyber-attacks, such as hacking or data theft [14]. Consider an AI system developed to monitor patient heart rates and predict heart failure. If patient data is not properly secured, a breach could expose not just the heart rate data but potentially other linked sensitive information like patient identities and addresses. Implementing encryption and access controls, along with ensuring legal compliance for data use, are essential steps in safeguarding this data [5].

To reduce the risks associated with AI in healthcare, particularly those concerning data privacy and security, a structured approach to securing sensitive information is essential. Implementing diverse security measures ensures the protection of patient data and enhancing trust in AI systems [1].

Data encryption serves as the first line of defence, securing data both in storage and in transit. By encrypting the data, healthcare organizations ensure that even if unauthorized parties access the information, it remains unreadable and protected. This step is crucial given the sensitive nature of health records and these records being vulnerable to cyber threats [14].

Data anonymization addresses privacy concerns directly associated with the use of real patient data in training AI models. By removing personally identifiable information from data sets before they are utilized in AI training, the risk of privacy breaches is significantly reduced. Anonymization ensures that even if data is somehow compromised, the information cannot be traced back to individuals, therefore safeguarding patient confidentiality [15].

Access control further strengthens security by ensuring that only authorized personnel have access to sensitive data. Strict authentication procedures, such as multi-factor authentication, are crucial in maintaining the integrity of data access. Limiting data access to individuals directly involved in the AI project minimizes potential exposure and reduces the risk of internal breaches [4].

To test the security measures, regular audits are important for maintaining system integrity over time. By conducting regular security audits and penetration testing, organizations can proactively identify, fix or improve vulnerabilities in their AI systems [4]. These audits help in early detection of potential security loopholes that could be exploited by malicious entities. Complying with legal frameworks such

as the General Data Protection Regulation (GDPR) in the European Union, which mandates strict data protection and privacy norms, is crucial [16] [23]. These laws also require that individuals provide informed consent for their data to be used, ensuring they are aware of how their information will be used and the associated risks.

B. 4.2 Bias and Fairness

Sallam discusses the potential for bias in AI algorithms, which can arise from biased training data or flawed algorithm design [2]. This raises ethical concerns about fairness and equality in AI-driven medical research outcomes, as biases can lead to inequality in diagnosis, treatment recommendations, and patient care [9].

AI algorithms learn patterns and make predictions based on the data they are trained on. If this training data is biased or unrepresentative of the diverse population it aims to serve, the AI model may unintentionally maintain or increase existing biases [9]. For example, if historical medical data primarily includes information from certain demographic groups, the AI model may struggle to accurately diagnose or recommend treatments for individuals from underrepresented groups [9].

Another example of bias in AI-driven medical systems is found in algorithms used for diagnosing skin cancer. If the training data primarily includes images of skin diseases from lighter-skinned individuals, the AI may be less accurate when diagnosing conditions on darker skin tones. This imbalance can result in misdiagnoses or delayed treatment for individuals with darker skin, maintaining healthcare inequalities. This kind of bias not only affects the fairness of medical services but also highlights the critical need for diverse datasets in training AI models to ensure equitable healthcare outcomes across different racial and ethnic groups [9].

The design and development of AI algorithms themselves can introduce biases. This can happen due to factors such as the choice of features to include in the model, the way in which data is weighted or processed, or the underlying assumptions made during algorithm development [17]. Even unintentional biases in the design of the algorithm can lead to unfair or discriminatory outcomes. Addressing bias and promoting fairness in AI-driven medical research is essential for ensuring equitable healthcare outcomes for all individuals. This requires careful attention to the sources of bias in AI algorithms, transparent algorithm design and implementation, and ongoing efforts to monitor and reduce biases in healthcare AI systems [12].

C. 4.3 Transparency and Explainability

Weinert et al. highlight how the complexity and non-transparent character of AI algorithms can lead to ethical issues [3]. For medical research to be ethical, it is important that AI models are clear and their decisions understandable. This clarity builds trust with doctors, researchers, and patients. It also ensures that people can hold these systems accountable and understand how they make their decisions.

1) **4.3.1 Importance of Transparency and Explainability in Medical AI:** Weinert et al. stresses that using AI in medical research and practice must be transparent and easy to understand [3]. It is important for AI models to be accurate and also clear about how they work. This transparency helps clinicians, researchers, and patients trust and understand AI decisions. This transparency is crucial for several reasons:

- **Trust:** AI systems must be trustworthy to be effectively integrated into healthcare settings. Trust is established when both the processes and outcomes of AI applications are transparent and understandable to all stakeholders [14].
- **Informed Consent:** Ethical research practices require that participants understand the role and function of AI in their care, ensuring that consent is truly informed [8].
- **Accountability:** Clear mechanisms for explaining AI decisions are essential for attributing responsibility, particularly when errors occur, or outcomes are unfavourable [8].
- **Fairness:** Explainability supports efforts to identify and mitigate biases within AI systems, promoting fairness in patient care and treatment outcomes [9].

2) **4.3.2 Strategies for Enhancing Transparency and Explainability:** To achieve transparency and explainability in medical AI, these are some examples of strategies that can be employed:

Development of Explainable AI Models: Layer-wise Relevance Propagation (LRP) and SHapley Additive exPlanations (SHAP) are examples of techniques that help to explain how AI makes decisions. LRP shows the impact of each input on the final decision by tracking how much each contributes as it moves through the AI's layers [18]. SHAP calculates how much each feature in the data affects the prediction by comparing it to an average prediction. Both methods make it easier for people who aren't AI experts to understand why an AI system made a specific decision [19].

- **Standardization of Documentation:** Detailed and standardized documentation of AI technologies should be mandated, describing the AI's design, functionality, decision-making process, and limitations. This documentation should be accessible to all stakeholders involved in healthcare delivery and research [8].
- **Regulatory Oversight:** Regulatory frameworks must be established to oversee the deployment of AI in healthcare, ensuring these technologies meet predefined standards of transparency and accountability before they are used in clinical settings.
- **Stakeholder Engagement:** Engaging with a diverse range of stakeholders, including patients, healthcare providers, and the public, is important for aligning AI systems with user needs and ethical standards. This engagement also contributes in educating stakeholders about the benefits and limitations of AI in healthcare [1] [2].
- **Education and Training:** Education and training programs

for healthcare professionals regarding the ethical use of AI are essential. These programs should cover the interpretation of AI decisions, understanding model limitations, and the ethical implications of AI technology [20].

In conclusion, as AI becomes more common in healthcare, being clear and understandable is very important. Transparency and explainability build trust, increase accountability, and make AI in healthcare more effective and fairer. Following these ethical guidelines, as emphasized by Weinert et al., is key to responsibly using AI in medical practices.

D. 4.4 Informed consent in AI-assisted Medical Research

Informed consent is the process of ensuring that a person fully understands and agrees to participate in a medical treatment or research after being informed of all the risks, benefits, and alternatives [5]. Hogg et al. discuss the importance of informed consent in AI-assisted medical research. They emphasize that patients need to be fully aware of how AI is used in their care, along with its risks and benefits, so they can make well-informed decisions about participating [5].

1) **4.4.1 Challenges in Ensuring Informed Consent with AI:** Ensuring informed consent with AI in healthcare is challenging due to three main issues. First, AI systems are complex, making it hard for patients to understand how their health decisions are made. Second, AI can change and learn over time, which might alter how it uses data after patients have already agreed to its use. Finally, AI uses large amounts of personal data, and patients need to know how their information is handled, who can access it, and how it is protected [10]. These factors complicate how patients give their informed consent.

2) **4.4.2 Practical Approaches to Enhance Informed Consent:** To maintain ethical standards in AI-assisted medical research, it is important to explain AI in simple and clear terms. This means providing clear, easy-to-understand information about how AI functions, what it is used for in medical treatment or research, and what results it aims to achieve. Using visual aids and simple comparisons can make these complex ideas easier to grasp [5].

Secondly, consent should be an ongoing process, not just a one-time agreement. Since AI technology can change and improve, it is important to keep patients updated and involved through regular discussions. The goal of this is to ensure that patients are well-informed and continue to agree to the use of AI as it evolves. It is also crucial to clearly explain how patient data is managed, protected, and accessed, including their rights to withdraw their data. Being open about sharing data with third parties and why it is shared is also important to keep patients' trust [14].

Lastly, it is important to involve a diverse group of people, including ethicists, legal experts, patients, and healthcare providers, in creating and improving consent processes.

Their different viewpoints contribute to making sure that the consent forms and procedures are ethical, effective, and meet legal and patient expectations. This approach enables to responsibly incorporate AI into medical research and practice, guided by ethical standards [5].

E. 4.5 Autonomy and the Patient-Physician Relationship

The use of AI in medical diagnostics and treatments is changing healthcare. While AI offers many benefits, like better diagnosis accuracy and tailored treatment plans, Gundersen and Bærøe point out concerns about its effect on the relationship between patients and doctors [4]. The main ethical issue is making sure that AI support patient independence while still keeping the important human aspects like empathy, understanding, and personalized care that are key to the patient-doctor relationship [4].

1) 4.5.1 The Impact of AI on Patient Autonomy:

Gundersen and Bærøe raise concerns about how AI in healthcare could challenge patient autonomy [4]. They point out that AI could dominate medical decision-making, potentially overshadowing doctors' insights and patients' preferences. The complexity of AI may also make it hard for patients to fully understand the treatments or diagnostics suggested, which is crucial for informed consent. Additionally, because AI uses a lot of personal data, there are worries about data privacy and who controls this information. These issues highlight the need to carefully manage AI's role in healthcare to protect essential ethical principles like patient autonomy [13].

2) 4.5.2 Preserving the Patient-Physician Relationship:

Gundersen and Bærøe emphasize that AI should support, not undermine, the human elements critical to the patient-physician relationship [4]. They propose several strategies to ensure this balance is maintained. First, AI should be viewed as a complementary tool that enhances, rather than replaces, the clinical judgment of a doctor and patient interactions. This means designing AI tools that support rather than reduce these key aspects. Second, doctors need thorough training to understand AI tools well and use them in ways that do not compromise their ability to connect with patients. Lastly, there should be specific regulatory frameworks in place to ensure that AI applications in healthcare promote patient autonomy and maintain the integrity of the patient-physician relationship. These steps are crucial for integrating AI into healthcare without losing the essential personal touch that defines good medical practice [13].

3) 4.5.3 Strategies for Enhancing Autonomy in AI-Integrated Care: To ensure that AI supports patient autonomy while integrating into medical diagnostics and treatment, Gundersen and Bærøe recommend several key strategies. Firstly, transparency is crucial, making AI algorithms more transparent ensures both patients and physicians understand the basis of

AI's recommendations, leading to better-informed decisions [4]. Secondly, patient engagement is vital, involving patients in the development and implementation of AI systems ensures that these technologies align with their needs and values. Lastly, ethical design is necessary. AI systems should be designed with a focus on ethical implications, particularly regarding autonomy and consent. This includes incorporating mechanisms that allow patients to control their data and opt out of AI-driven care when appropriate.

As AI becomes more common in healthcare, keeping the important human aspects of the doctor-patient relationship becomes more difficult. Gundersen and Bærøe emphasize the need for AI technologies that support, rather than reduce, patient independence and personalized care [4]. By adopting ethical approaches to AI use, the medical community can preserve these key values, ensuring that AI helps rather than hinders healthcare while upholding the basic principles of medical ethics [4].

F. 4.6. Accountability and Liability in AI-Assisted Medical Research

When AI systems in medical research make mistakes or cause harm, it raises difficult questions about who is responsible. The widespread involvement in creating and using AI makes it hard to pinpoint who is accountable. This complexity means we need new ethical rules to properly handle these issues of responsibility and liability. Gundersen and Bærøe discuss the ethical issues that come with these technological changes [4]. Giving a deeper understanding on how to determine responsibility and liability regarding use of AI in healthcare.

1) 4.6.1 Challenges in AI Accountability and Liability:

Establishing clear accountability and liability in AI-assisted medicine is challenging due to several key factors, as Gundersen and Bærøe highlight [3]. First, AI systems are often developed and used by a diverse group of people, including developers, data scientists, clinicians, and other healthcare providers. This makes it difficult to identify who is responsible when errors occur, as responsibility is shared among many. Secondly, the decision-making process in AI, driven by complex and sometimes non-transparent algorithms, is hard to trace, which complicates efforts to pinpoint the source of errors or failures. Lastly, AI systems can change and adapt on their own over time as they process new data, making it even harder to determine who is accountable for their actions. These factors collectively make it challenging to assign clear liability in AI-driven healthcare settings [4].

2) 4.6.2 Proposals for Ethical Frameworks: To address these ethical challenges, Gundersen and Bærøe suggest several strategies to improve accountability and clarify liability in AI-assisted medical research:

- **Transparent Reporting:** We need to set clear standards for openness in how AI is developed and used. This includes

thoroughly documenting how AI makes decisions and defining the roles of everyone involved.

- **Regulatory Oversight:** Create detailed regulations for AI in healthcare, focusing on proper testing and validation before these systems are used in clinics.
- **Ethical Auditing:** Conduct regular audits to check that AI systems meet ethical and legal standards, and to evaluate how AI decisions affect patient care.
- **Liability Insurance Models:** Develop new insurance options that cover risks specific to AI in healthcare, possibly including shared responsibility between AI developers and healthcare providers.

The challenges of accountability and liability in AI-assisted medical research are complex, requiring innovative legal and ethical solutions. Gundersen and Bærøe emphasize the need for strong collaboration across various sectors to create frameworks that align AI use with ethical and legal standards. This includes technologists, ethicists, healthcare professionals, and policymakers working together to ensure responsible AI integration in healthcare [4].

G. 4.7 SRQ1.1 What are the implications of Data Privacy and Consent in AI-Driven Data Collection and Analysis

The integration of AI in data collection and analysis presents significant ethical, legal, and operational challenges, particularly concerning data privacy and consent. Ethically, AI systems must respect individual privacy and secure informed consent, especially critical in healthcare settings. Legally, they must comply with stringent regulations like the GDPR in the European Union, which mandates explicit consent for data processing and stringent privacy safeguards to avoid penalties and maintain public trust [11].

From an operational perspective, the vast amount of data AI processes increases the risk of breaches. It is essential to protect data during both storage and transit through advanced security measures and constant monitoring. Additionally, AI systems need to be transparent and fair, avoiding biases that could affect outcomes and ensuring users can understand and verify how their data is used and decisions are made. This includes giving individuals control over their data, with clear options to withdraw consent and strict adherence to data minimization principles, collecting only what is necessary for specific purposes and clearly defining how long and for what reasons the data will be used. These measures collectively ensure that AI's integration into data handling respects both ethical norms and legal standards while safeguarding user autonomy and trust [17].

H. 4.8 SRQ1.2 How does the deployment of AI in medical research intersect with issues of bias and fairness?

As AI systems in healthcare use large datasets, it is crucial to address any biases in the data to prevent discrimination in patient care. The ethical concerns go beyond just data management to include the transparency of algorithms and the accountability of those who manage AI in healthcare [10].

Efforts to reduce these biases involve creating guidelines that focus on fairness, accountability, and transparency, developed by a mix of researchers, ethicists, and healthcare professionals. These experts stress the need for ethical AI that promotes equal healthcare outcomes. As technology advances, continuing discussions and updates to these ethical guidelines are essential to ensure AI improves healthcare delivery and does not increase existing inequalities [2] [9].

I. 4.9 SRQ1.3 What ethical challenges arise from the potential lack of transparency and explainability in AI models used in medical research?

AI in healthcare has revolutionized diagnosis and treatment but also brings significant ethical challenges, especially around transparency and explainability. AI systems can act like "black boxes," making it hard for healthcare professionals to see or explain how decisions are made. This lack of clarity can reduce trust and make it difficult to ensure informed consent from patients [13].

Additionally, this lack of transparency might increase existing healthcare inequalities. If AI is trained on data that lacks diversity, it might perform poorly for certain groups of patients, increasing rather than decreasing health disparities [12]. To address these issues, ethical guidelines and frameworks are being developed to make AI in healthcare more transparent, accountable, and fair. These efforts aim to maximize the benefits of AI while minimizing risks and maintaining the ethical standards of medical practice.

V. 5. EXISTING FRAMEWORKS AND GUIDELINES REGARDING ETHICS IN MEDICAL AI

A. 5.1 RQ2. What frameworks or guidelines have been proposed or implemented to address these ethical concerns?

This section will discuss some of the prominent institutions and governing bodies that are developing and implementing frameworks and guidelines regarding the use of AI in the medical field, along with some existing framework on how ethical AI should be outlined. The use of AI in healthcare raises important ethical, legal, and social issues. To address these concerns and ensure AI is used safely and ethically, various ethical frameworks and guidelines have been established. Some examples of existing governing or regulatory entities are the European Union Ethical Guidelines for Trustworthy AI and World Health Organization (WHO) guidelines [8] [11]. These frameworks stress the importance of making AI systems safe, fair, and transparent to align with the key goals of medical ethics and patient care. The aim is to ensure that AI in healthcare works effectively while respecting human dignity and maintaining public trust [7].

The European Union Ethical Guidelines for Trustworthy AI outline principles to ensure AI systems across EU member states are safe, ethical, and effective. These guidelines stress commitment to laws, respect for human rights, and robustness in operation [11]. Important aspects include enhancing human decision-making, ensuring AI is secure and reliable, protecting privacy, and maintaining transparency in AI operations. AI

must also be inclusive, fair, environmentally sustainable, and accountable for its actions. These principles aim to create AI that is beneficial and trustworthy, aligning with EU values and ethical standards.

B. 5.2 WHO Guidance on AI Ethics and Governance

The WHO has recently issued new guidelines focusing on the ethical use and governance of large multi-modal AI models (LMMs) in healthcare. These guidelines set clear standards to ensure that AI technologies not only enhance healthcare delivery but also adhere to ethical norms and protect patient safety. Specifically, the WHO guidance emphasizes the importance of transparency, accountability, and inclusivity in the development and deployment of AI systems [8].

The guidelines advocate for the continuous monitoring of AI applications to evaluate their impact on health outcomes and equity. They also call for robust privacy protections to safeguard patient data and stress the necessity of involving diverse stakeholders, including patients and public health experts, in the AI development process. This inclusive approach is designed to ensure that AI tools are developed with a deep understanding of the varied healthcare environments they will operate in, promoting better health outcomes and minimizing potential harms. By setting these standards, the WHO aims to guide the development of AI technologies in medicine to be both ethically in place and effective [8]. Box 1 shows a brief overview of WHO consensus ethical principles for use of AI for health.

Box 1. Brief overview of WHO consensus ethical principles for use of AI for health

- **Protect autonomy:** Humans should remain in control of health-care systems and medical decisions. Providers have the information necessary to use AI systems safely and effectively. People understand the role that AI systems play in their care. Data privacy and confidentiality are protected by valid informed consent through appropriate legal frameworks for data protection.
- **Promote human well-being, human safety and the public interest:** Designers of AI satisfy regulatory requirements for safety, accuracy and efficacy for well-defined uses or indications. Measures of quality control in practice and quality improvement in the use of AI over time should be available. AI is not used if it results in mental or physical harm that could be avoided by use of an alternative practice or approach.
- **Ensure transparency, “explainability” and intelligibility:** AI technologies should be intelligible or understandable to developers, medical professions, patients, users and regulators. Sufficient information is published or documented before the design or deployment of AI, and the information facilitates meaningful public consultation and debate on how the AI is designed and how it should or should not be used. AI is explainable according to the capacity of those to whom it is explained.
- **Foster responsibility and accountability** to ensure that AI is used under appropriate conditions and by appropriately trained people. Patients and clinicians evaluate development and deployment of AI. Regulatory principles are applied upstream and downstream of the algorithm by establishing points of human supervision. Appropriate mechanisms are available for questioning and for redress for individuals and groups that are adversely affected by decisions based on AI.
- **Ensure inclusiveness and equity:** AI is designed and shared to encourage the widest possible, appropriate, equitable use and access, irrespective of age, sex, gender identity, income, race, ethnicity, sexual orientation, ability or other characteristics. AI is available for use not only in high-income settings but also in low- and middle-income countries. AI does not encode biases to the disadvantage of identifiable groups. AI minimizes inevitable disparities in power. AI is monitored and evaluated to identify disproportionate effects on specific groups of people.
- **Promote AI that is responsive and sustainable:** AI technologies are consistent with the wider promotion of the sustainability of health systems, the environment and workplaces.

Fig. 3: Overview of WHO consensus ethical principles for use of AI for health [8]

C. 5.3 Systematic Frameworks for Ethical AI

Systematic frameworks for ethical AI offer a clear method for handling ethical issues during the design and use of AI systems in healthcare. These frameworks guide researchers and developers in spotting and managing potential ethical problems, from data privacy to fairness and transparency. They include steps like conducting detailed ethical evaluations, involving a variety of stakeholders in planning, and constantly monitoring AI applications to see how they affect patients and ensure they are fair.

The guidelines also emphasize strict data protection measures to keep patient information safe and the importance of designing AI tools that meet the diverse needs of different patient groups. Essentially, these frameworks try to ensure that AI technologies in healthcare are not only effective but also ethically responsible, leading to safer and more equitable healthcare solutions.

D. 5.4 SHIFT Framework

The SHIFT Framework outlines six key ethical principles for developing AI in healthcare: fairness, accountability, transparency, inclusivity, privacy, and security. This framework ensures that AI technologies are developed responsibly, aligning with societal values and healthcare ethics [21]. It emphasizes creating AI systems that are fair, transparent, inclusive, secure, and protective of privacy. This approach aids developers incorporate these ethical considerations into every stage of AI development, leading to more effective and ethical healthcare solutions. By following these frameworks, the medical community can use AI to enhance health outcomes while dealing with the ethical challenges that come with technological advances [21].

E. 5.6 SRQ2.1 What existing ethical frameworks and guidelines are being adapted or developed to address the unique challenges posed by AI in medical research?

Key guidelines like the Asilomar AI Principles focus on safety and aligning medical AI with human values. The European Union Ethical Guidelines ensure that AI in healthcare is legal, ethical, and robust, emphasizing patient autonomy, harm prevention, fairness, and clarity. Additionally, the WHO provides guidance on AI ethics, promoting transparency, inclusiveness, and accountability to benefit health globally [8] [11]. Laws like Health Insurance Portability and Accountability Act (HIPAA) and GDPR adapt to AI’s complexities to ensure strong data protection and patient privacy [22] [23]. The IEEE Standards encourage AI tool development based on clinical needs and ethical standards, supporting teamwork among engineers, clinicians, and ethicists [4]. Together, these frameworks help address the ethical issues AI brings to medical research, ensuring patient safety and integrity.

F. 5.7 SRQ2.2 How are regulatory bodies responding to the rapid advancement and integration of AI technologies in medical research in terms of policy making and governance?

Regulatory bodies are working to keep up with the rapid growth of AI in medical research by developing policies

and governance frameworks. Recognizing the need for safety, efficacy, and ethics, agencies like the FDA in the United States and the European Medicines Agency in Europe have begun issuing guidelines for using AI in healthcare [11] [24]. These guidelines ensure that AI systems are transparent, explainable, fair, and comply with medical device regulations.

Additionally, rules like the GDPR in Europe set strict privacy guidelines for AI applications that manage sensitive health data. These regulations aim to protect patient privacy, ensure data accuracy, and maintain trust in the healthcare system [23].

Global organizations like the WHO are also providing international guidance to standardize AI practices worldwide. This includes promoting ethical standards, improving health access equity, and ensuring that AI benefits all population segments. Through these efforts, regulatory bodies play a crucial role in creating a safe and ethical environment for AI in medical research [8].

VI. 6. CONCLUSION

The integration of AI into medical research and healthcare promises significant advancements in diagnostics, patient care, and treatment customization. However, the fast development of AI also brings substantial ethical challenges that require careful management to ensure responsible use. This review has addressed various important ethical considerations surrounding AI applications in healthcare, such as data privacy, security, fairness, transparency, and informed consent.

It is clear from our exploration that following established ethical frameworks and guidelines is essential. Key documents like the European Union Ethical Guidelines for Trustworthy AI, and WHO AI ethics guidelines play a critical role in guiding the ethical use of AI in healthcare. These guidelines help ensure that AI not only meets legal requirements but also aligns with societal ethical expectations, thus building trust and promoting fair health outcomes.

AI holds huge potential for improving healthcare but must be implemented with a strong commitment to ethical practices. Ongoing discussions, collaborative efforts across disciplines, and strict commitment to ethical standards are important for benefiting from AI responsibly. Future efforts should focus on refining ethical guidelines, increasing stakeholder engagement, and encouraging international cooperation to navigate the evolving ethical landscape of AI in healthcare.

However, it is important to stress that most of the literature covered in this paper offers guidelines but cannot combat or prevent the misuse of AI or the use of AI outside the boundaries defined by ethical guidelines. The ethical guidelines discussed in this paper will need to be supplemented with tools that can enforce these guidelines and implement data policies that are checked before, during, and after data processing. Only through such comprehensive measures can we ensure that AI technologies are used ethically and effectively in healthcare, safeguarding patient welfare and maintaining high ethical standards.

VII. 7. FUTURE WORK

As AI technology advances and becomes more common in healthcare, it is important to keep updating the ethical guidelines to ensure it is used responsibly. Future efforts should focus on refining these ethical guidelines to keep up with new technologies and societal changes. This means bringing in a diverse group of stakeholders, including patients, healthcare providers, ethicists, and technologists, to make sure AI solutions are fair and meet the needs of everyone involved.

An interesting area for future research is data policy enforcement. While this paper provides an overview of guidelines and frameworks regarding AI ethics in medical research, the next step involves the enforcement of these policies to prevent the misuse of data before, during, and after data processing. This would ensure that AI applications adhere to ethical standards throughout their lifecycle. The Multiscale Networked group of the University of Amsterdam is developing techniques and frameworks to help address the data privacy enforcement [26] [27] [28] [29].

There is also a crucial need to make AI systems effective, transparent, and easy to understand for everyone. This transparency will build trust and make AI more usable in clinical settings. It is important to work together internationally to align AI regulations and standards around the world. This collaboration will help address global challenges and ensure consistent ethical practices across different regions.

Additionally, carrying out long-term studies to examine the effects of AI on patient outcomes will give important insights that can improve AI applications in healthcare. By focusing on these areas, the medical community can use AI responsibly, ensuring it improves patient care while meeting the highest ethical standards.

REFERENCES

- [1] Al Kuwaiti, A., Nazer, K., Al-Reedy, A., Al-Shehri, S., and Alrowaili, Z. (2023). A review of the role of artificial intelligence in enhancing patient safety in healthcare. *Journal of Personalized Medicine*, 13(6), 951. <https://doi.org/10.3390/jpm13060951>
- [2] Sallam, M. (2023). ChatGPT utility in healthcare education, research, and clinical practice: A scoping review. *Healthcare (Switzerland)*, 11(6), 887. <https://doi.org/10.3390/healthcare11060887>
- [3] Weinert, L., Klass, M., Schneider, G., and Heinze, O. (2022). Exploring stakeholder requirements to enable the ethical use of AI in medical imaging. *JMIR Research Protocols*, 11(12), e42208. <https://doi.org/10.2196/42208>
- [4] Gundersen, T., and Bærøe, K. (2022). The future ethics of artificial intelligence in health care: A comprehensive overview. *Science and Engineering Ethics*, 28(2), 17. <https://doi.org/10.1007/s11948-022-00350-5>
- [5] Hogg, H.D.J., Brittain, K., Teare, D., Talks, J., and Lotery, A. (2023). Safety and efficacy of an artificial intelligence system in a real-world clinical setting. *BMJ Open*, 13(2). <https://doi.org/10.1136/bmjopen-2022-069443>
- [6] Zhang, J., and Zhang, Z. M. (2023). Ethics and governance of trustworthy medical artificial intelligence. *BMC Medical Informatics and Decision Making*, 23, 7. <https://doi.org/10.1186/s12911-023-02029-2>
- [7] Crossnohere, N. L., Elsaid, M., Paskett, J., Bose-Brill, S., and Bridges, J. F. P. (2022). Guidelines for artificial intelligence in medicine: Literature review and content analysis of frameworks. *Journal of Medical Internet Research*, 24(8), e37336. <https://doi.org/10.2196/37336>
- [8] Heinrichs, B., and Eickhoff, S. B. (2020). Your evidence? Machine learning algorithms for medical diagnosis and prediction. *Human Brain Mapping*, 41(6), 1435–1444. <https://doi.org/10.1002/hbm.24912>

- [9] Xafis, V., Schaefer, G. O., Labude, M. K., et al. (2019). An ethics framework for big data in health and research. *Asian Bioethics Review*, 11, 227–254. <https://doi.org/10.1007/s41649-019-00093-5>
- [10] Geis, J. R., Brady, A. P., Wu, C. C., Spencer, J., Ranschaert, E., Jaremko, J. L., Langer, S. G., Kitts, A. B., Birch, J., Shields, W. F., van den Hoven van Genderen, R., Kotter, E., Gichoya, J. W., Cook, T. S., Morgan, M. B., Tang, A., Safdar, N. M., and Kohli, M. (2019). Ethics of artificial intelligence in radiology: Summary of the joint European and North American multisociety statement. *Radiology*, 293(2), 436–440. <https://doi.org/10.1148/radiol.2019191586>
- [11] Goisaufl, M., and Cano Abadía, M. (2022). Ethics of AI in radiology: A review of ethical and societal implications. *Frontiers in Big Data*, 5, 850383. <https://doi.org/10.3389/fdata.2022.850383>
- [12] Fiske, A., Henningsen, P., and Buyx, A. (2019). Your robot therapist will see you now: Ethical implications of embodied artificial intelligence in psychiatry, psychology, and psychotherapy. *Journal of Medical Internet Research*, 21(5), e13216. <https://doi.org/10.1198/jmir.2019191586>
- [13] Rosemann, A., and Zhang, X. (2022). Exploring the social, ethical, legal, and responsibility dimensions of artificial intelligence for health – a new column in Intelligent Medicine. *Intelligent Medicine*, 2(2).
- [14] Car, J., Sheikh, A., Wicks, P. et al. Beyond the hype of big data and artificial intelligence: building foundations for knowledge and wisdom. *BMC Med* 17, 143 (2019).
- [15] Cope, H., Willis, C. R. G., MacKay, M. J., Rutter, L. A., Toh, L. S., Williams, P. M., Herranz, R., Borg, J., Bezdán, D., Giacomello, S., Muratani, M., Mason, C. E., Etheridge, T., and Szewczyk, N. J. (2022). Routine omics collection is a golden opportunity for European human research in space and analog environments. *Patterns* (New York, N.Y.), 3(10), 100550. <https://doi.org/10.1016/j.patter.2022.100550>
- [16] Garcha, I., and Phillips, S. P. (2023). Social bias in artificial intelligence algorithms designed to improve cardiovascular risk assessment relative to the Framingham Risk Score: a protocol for a systematic review. *BMJ open*, 13(5), e067638. <https://doi.org/10.1136/bmjopen-2022-067638>
- [17] Bach, S., Binder, A., Montavon, G., Klauschen, F., Müller, K. R., and Samek, W. (2015). On Pixel-Wise Explanations for Non-Linear Classifier Decisions by Layer-Wise Relevance Propagation. *PLoS one*, 10(7), e0130140.
- [18] Lundberg, S. M., and Lee, S. I. (2017). A Unified Approach to Interpreting Model Predictions. In *Advances in Neural Information Processing Systems* (pp. 4765–4774).
- [19] Weidener, L., and Fischer, M. (2023). Artificial Intelligence Teaching as Part of Medical Education: Qualitative Analysis of Expert Interviews. *JMIR medical education*, 9, e46428. <https://doi.org/10.2196/46428>
- [20] Haytham Siala, Yichuan Wang, SHIFTing artificial intelligence to be responsible in healthcare: A systematic review, *Social Science and Medicine*, Volume 296, 2022,
- [21] Baumer, D., Earp, J. B., and Payton, F. C. (2000). Privacy of medical records: IT implications of HIPAA. *ACM SIGCAS Computers and Society*, 30(4), 40–47.
- [22] Hansen, J., Wilson, P., Verhoeven, E., Kroneman, M., Kirwan, M., Verheij, R., and van Veen, E. B. (2021). Assessment of the EU Member States’ rules on health data in the light of GDPR.
- [23] Harvey, H. B., and Gowda, V. (2020). How the FDA regulates AI. *Academic radiology*, 27(1), 58–61.
- [24] World Health Organization. (2024). Ethics and governance of artificial intelligence for health: Guidance on large multi-modal models. <https://doi.org/978-92-4-008475-9>
- [25] <https://github.com/NLeSC/litstudy>
- [26] C. A. Esterhuysen, et al. "Exploring the Enforcement of Private, Dynamic Policies on Medical Workflow Execution," 2022 IEEE 18th International Conference on e-Science (e-Science), Salt Lake City, UT, USA, 2022, pp. 481–486, <https://doi.org/10.1109/eScience55777.2022.000>
- [27] Kassem, J. et al. "EPI framework: Approach for traffic redirection through containerized network functions" 17th International Conference on eScience: 2021. doi: <https://doi.org/10.1109/eScience51609.2021.00018>
- [28] Amiri, Saba, et al. "On the impact of non-IID data on the performance and fairness of differentially private federated learning." 2022 52nd Annual IEEE/IFIP Int. Conference on Dependable Systems and Networks Workshops (DSN-W). IEEE, 2022.
- [29] Amiri, Saba, et al. "Differential Privacy vs Detecting Copyright Infringement: A Case Study with Normalizing Flows" GenLaw '23 in person in Honolulu, Hawai'i