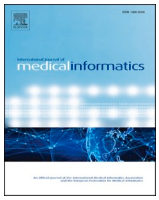
International Journal of Medical Informatics 195 (2025) 105780

Contents lists available at ScienceDirect

International Journal of Medical Informatics

journal homepage: www.elsevier.com/locate/ijmedinf



Review article

Towards secure and trusted AI in healthcare: A systematic review of emerging innovations and ethical challenges

Muhammad Mohsin Khan [a,*], Noman Shah [e], Nissar Shaikh [d], Abdulnasser Thabet [a], Talal alrabayah [a], Sirajeddin Belkhair [a,b,c]

[a] Neurosurgery Department, Hamad General Hospital, Qatar
[b] Department of Clinical Academic Sciences, College of Medicine, Qatar University, Doha, Qatar
[c] Department of Neurological Sciences, Weill Cornell Medicine, Doha, Qatar
[d] Surgical Intensive Care Unit, Hamad General Hospital, Qatar
[e] Neurosurgery Department, Abbottabad Medical Complex, Pakistan

ARTICLE INFO

Keywords:
Artificial Intelligence (AI)
Healthcare
Trust
Safety
Ethics
Explainability
Transparency
Patient safety

ABSTRACT

Introduction: Artificial Intelligence is in the phase of health care, with transformative innovations in diagnostics, personalized treatment, and operational efficiency. While having potential, critical challenges are apparent in areas of safety, trust, security, and ethical governance. The development of these challenges is important for promoting the responsible adoption of AI technologies into healthcare systems.

Methods: This systematic review of studies published between 2010 and 2023 addressed the applications of AI in healthcare and their implications for safety, transparency, and ethics. A comprehensive search was performed in PubMed, IEEE Xplore, Scopus, and Google Scholar. Those studies that met the inclusion criteria provided empirical evidence, theoretical insights, or systematic evaluations addressing trust, security, and ethical considerations.

Results: The analysis brought out both the innovative technologies and the continued challenges. Explainable AI (XAI) emerged as one of the significant developments. It made it possible for healthcare professionals to understand AI-driven recommendations, by this means increasing transparency and trust. Still, challenges in adversarial attacks, algorithmic bias, and variable regulatory frameworks remain strong. According to several studies, more than 60 % of healthcare professionals have expressed their hesitation in adopting AI systems due to a lack of transparency and fear of data insecurity. Moreover, the 2024 WotNot data breach uncovered weaknesses in AI technologies and highlighted the dire requirement for robust cybersecurity.

Discussion: Full understanding of the potential of AI will be possible only with putting into practice of ethical and technical maintains in healthcare systems. Effective strategies would include integrating bias mitigation methods, strengthening cybersecurity protocols to prevent breaches. Also by adopting interdisciplinary collaboration with the goal of forming transparent regulatory guidelines. These are very important steps toward earning trust and ensuring that AI systems are safe, reliable, and fair.

Conclusion: AI can bring transformative opportunities to improve healthcare outcomes, but successful implementation will depend on overcoming the challenges of trust, security, and ethics. Future research should focus on testing these technologies in multiple real-world settings, enhance their scalability, and fine-tune regulations to facilitate accountability. Only by combining technological innovations with ethical principles and strong governance can AI reshape healthcare, ensuring at the same time safety and trustworthiness.

1. Introduction

Artificial Intelligence (AI) is quickly transforming the landscape of healthcare, introducing ground-breaking methodologies for diagnosis and treatment besides patient management. By handling massive amounts of data, AI systems can classify patterns and make forecasts that have the potential to progress clinical outcomes and operational proficiency. On the other hand, together with these advancements,

* Corresponding author.
E-mail address: mmkyousafzai@gmail.com (M. Mohsin Khan).

https://doi.org/10.1016/j.ijmedinf.2024.105780
Received 13 November 2024; Received in revised form 21 December 2024; Accepted 27 December 2024
Available online 30 December 2024

substantial concerns regarding safety, trust, security, and ethical implications have developed. Addressing these issues is authoritative for AI to fulfill its promise in healthcare while safeguarding patient safety and maintaining public trust in the technology.

AI technologies, for the most part machine learning (ML) and deep learning algorithms, have established considerable efficiency in various healthcare applications. For example, AI has been employed to envisage patient health issues, evaluate medical imaging, and support surgical procedures, in so doing augmenting the capability of healthcare providers to deliver modified treatment and improve patient outcomes. A systematic review by Li et al. highlights the transformative role of AI in clinical practice, highlighting its potential to simplify precision medicine and improve patient care passageways through data-driven understandings and automation [1].

In the face of the potential benefits, the incorporation of AI in healthcare is fraught with risks. The intricacy of AI algorithms, frequently stated as "black boxes," stances challenges in understanding how decisions are made. This lack of transparency can lead to disbelief among healthcare providers and patients, as they may be uncertain to be dependent on AI-driven endorsements devoid of grasping the underlying rationale. A study by the HITRUST Alliance accentuates that trust is a precarious factor for fruitful AI adoption in healthcare, perceiving that over 60 % of patients express skepticism as regards of AI technologies in line for concerns about data privacy and algorithmic biases [2].

Safeguarding the safety of AI systems in healthcare is fundamental to avoid potential harm to patients. Robust AI systems need to be designed to execute consistently under innumerable conditions and repel adversarial attacks. Proper verification processes, which statistically prove the correctness of algorithms, along with wide-ranging testing, are indispensable before deploying AI in clinical settings. In addition, AI systems must be armed to handle failures with poise, categorizing impending points of failure and executing safety measures to alleviate risks. For illustration, in clinical decision support systems, AI should be proficient of flagging suspicions and allowing for human intervention when required, in this manner enhancing patient safety [3].

Confidence in AI is further strengthened through the development of explainable AI (XAI), which purposes to make AI processes clear as crystal and understandable. Explainability is vital for healthcare professionals to be familiar with the underlying principle behind AI-generated diagnoses or treatment suggestions. Techniques for example, attention mechanisms in neural networks, feature importance scores, and model-agnostic methods like LIME (Local Interpretable Model-agnostic Explanations) are being explored to increase the transparency of AI systems. By providing comprehensions into their workings and clear explanations for their estimates, XAI can raise trust among stakeholders and simplify the responsible use of AI in clinical practice [4].

Transparency is also attained through actual documentation and open communication on the subject of the capabilities and limitations of AI technologies. Forming realistic expectations about what AI can and cannot do is necessary for constructing trust amongst patients and healthcare providers. A broad understanding of AI's role in healthcare can lessen concerns and uphold a collaborative environment wherever AI augments human decision-making relatively than replacing it.

Safety is a paramount concern for AI systems in healthcare, assumed the delicate nature of medical data. Following to principles of confidentiality, integrity, and availability is dire for upholding patient trust and ensuring obedience with regulations such as the Health Insurance Portability and Accountability Act (HIPAA) in the United States. Robust security measures, including data encryption, access control, and regular security audits, are needed to protect AI systems from cyber threats. Moreover, ethical considerations nearby patient data usage for AI training must be addressed. Techniques similar to federated learning allow for AI model training on decentralized data without compromising patient privacy, by this means ensuring responsible AI development [2,4].

Ethical considerations are important to the placement of AI in healthcare, guaranteeing that it benefits all patients without introducing new risks or worsening existing inequalities. Bias in AI can arise from biased training data, leading to unequal treatment of different patient groups. Addressing fairness and accountability in AI systems necessitates the use of various and representative datasets, as well as continuing monitoring for biased outcomes. Ethical frameworks and guidelines from professional organizations and regulatory bodies can guide the liable development and use of AI in healthcare, encouraging equitable access to AI-driven innovations [1,3].

Accountability is indispensable for addressing the ethical implications of AI decisions. Clear policies should be made for delineating the roles and responsibilities of AI developers, healthcare providers, and other stakeholders, which is obligatory to ensure accountability in AI deployment. By placing AI as a tool to support human decision-making rather than replace it, healthcare organizations can sustain ethical standards and nurture a culture of responsibility in AI usage.

Despite the burgeoning body of literature on artificial intelligence in healthcare, there remains a profound lack of in-depth research into several of the key ethical issues—above all, algorithmic bias, data confidentiality, and transparency. Most of the current literature explores these individually, without an overall framework that makes clear how they are intertwined and together elevate risks to patient safety and trust. Moreover, current literature has poorly explored new innovations, such as Explainable AI (XAI) and federated learning, which hold potential solutions to these challenges. This review paper attempts to fill these gaps by reviewing recent developments of AI, identifying ethical and practical challenges, and providing actionable frameworks for secure and trustworthy AI systems in the healthcare sector.

This systematic review aims to provide a comprehensive evaluation of artificial intelligence innovations that address ethical and security issues. By way of analyzing existing gaps in the frameworks designed to ensure the virtues of trust, safety, and transparency in healthcare AI applications, this study brings forward interdisciplinary recommendations for practitioners, policymakers, and researchers. This way, it contributes to the comprehensive discussion of responsible artificial intelligence in the health sector and tries to lay the foundation for future empirical studies and policy-making. Addressing these critical issues, the review serves not only to enhance understanding but also to encourage the development of AI technologies that are safe, reliable, and trusted by health professionals and patients themselves.

In conclusion, despite the fact that AI presents substantial opportunities for advancing healthcare, speaking the associated challenges of safety, trust, security, and ethics is critical for its efficacious integration. By prioritizing explainability, robust security measures, and ethical frameworks, stakeholders can generate a healthcare environment that leverages AI reliably and effectively, at the end of the day benefiting patients and enhancing the quality of care.

2. Materials and methods

2.1. Study design

This systematic review was conducted to assess the state of research and development in technologies aimed at improving AI safety, trust, security, and responsible use in healthcare. The review followed the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines to ensure a comprehensive and transparent evaluation of the existing literature.

2.2. Search strategy

A comprehensive literature search was conducted across multiple databases including PubMed, IEEE Xplore, Scopus, and Google Scholar. The search string used included a combination of keywords and Boolean operators to capture relevant studies:

((“Artificial Intelligence” OR “AI” OR “Machine Learning” OR “Deep Learning” OR “Neural Networks”) AND (“Safety” OR “Trust” OR “Security” OR “Responsible Use” OR “Ethics” OR “Fairness” OR “Bias Mitigation” OR “Transparency” OR “Explainability” OR “Robustness” OR “Reliability”) AND (“Healthcare” OR “Medical” OR “Clinical” OR “Health Informatics” OR “Health Systems” OR “Patient Care”)).

The search was limited to studies published in English from January 2010 to December 2023. Reference lists of identified articles were also screened for additional relevant studies.

2.3. Inclusion and exclusion criteria

Studies were included in the review if they met the following criteria:

1. Focused on AI technologies used in healthcare.
2. Addressed aspects of safety, trust, security, or responsible use.
3. Provided empirical evidence, theoretical frameworks, or comprehensive reviews.
4. Published in peer-reviewed journals.

Studies were excluded from the review if they did not met the following criteria:

1. Studies not directly related to healthcare applications.
2. Articles that did not address AI safety, trust, security, or ethical considerations.
3. Non-peer-reviewed articles, opinion pieces, and editorials.

2.4. Quality assessment

The quality of the included studies was assessed using the Joanna Briggs Institute (JBI) Critical Appraisal Checklist for Systematic Reviews and Research Syntheses. This tool evaluates the methodological rigor and relevance of studies based on criteria such as clarity of research questions, appropriateness of study design, and robustness of data analysis.

Each study was scored on a scale of 0 to 10, with higher scores

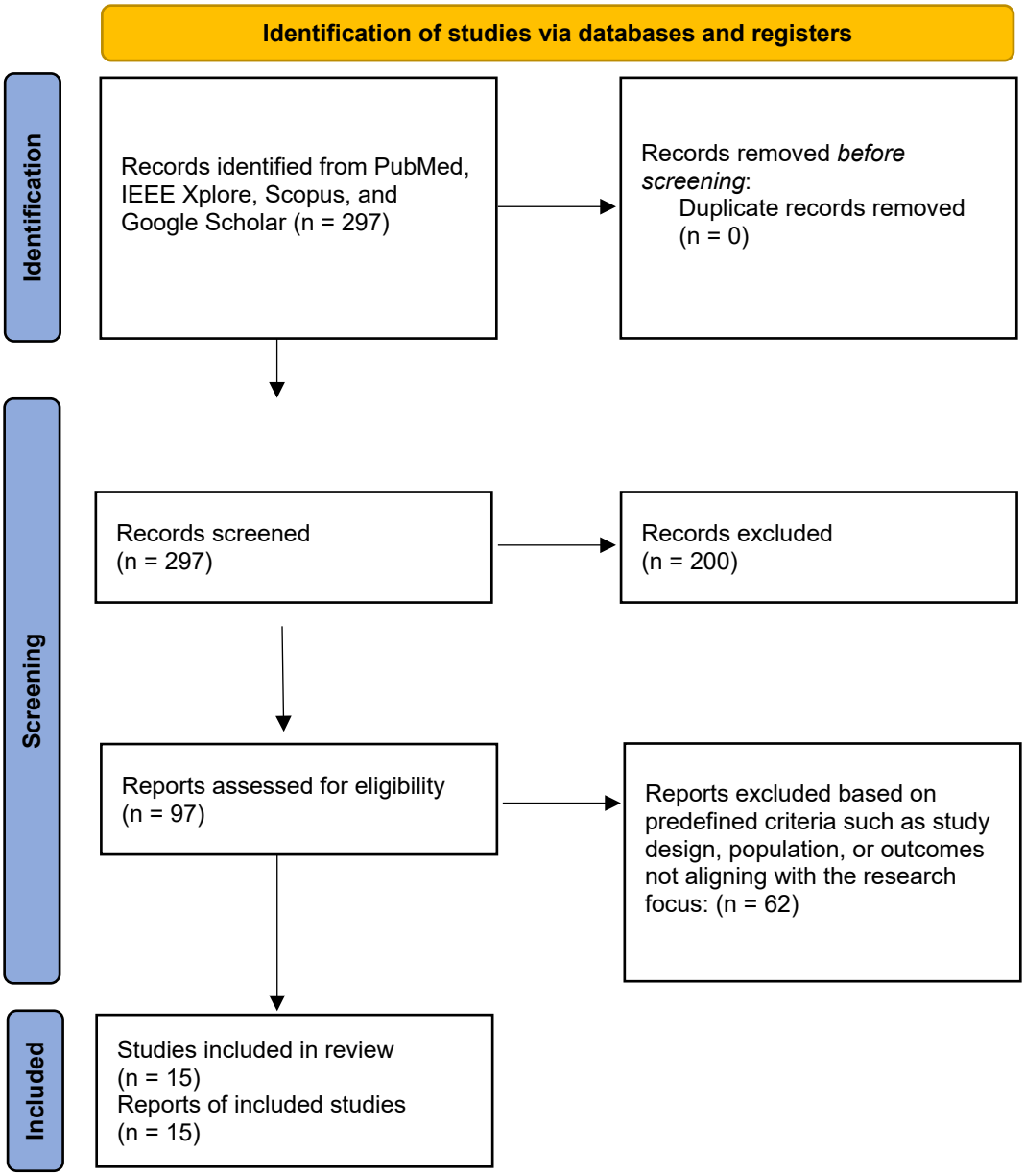
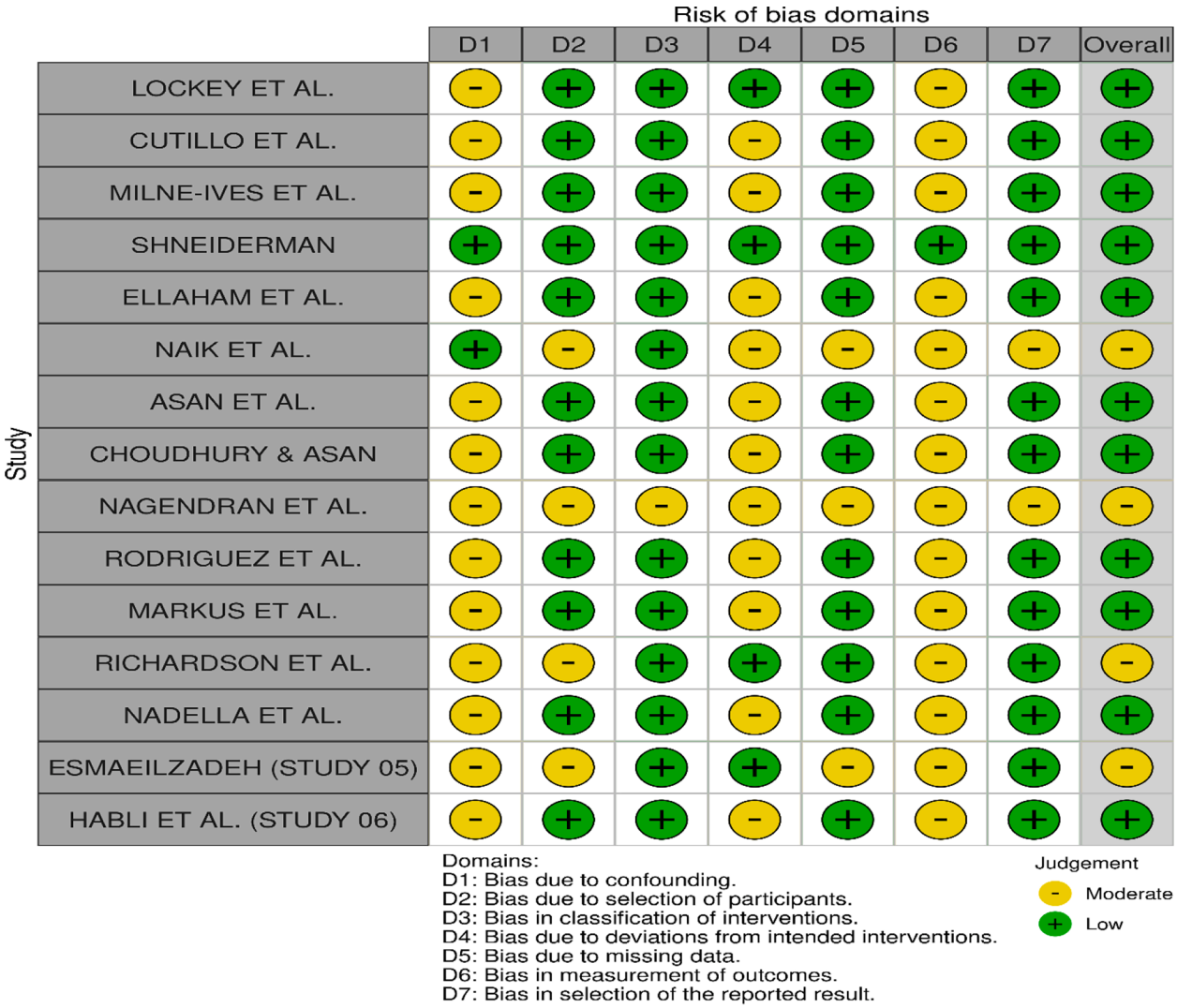


Fig. 1. PRISMA flowchart.

indicating better quality. Studies scoring below 5 were considered of lower quality and were subjected to sensitivity analysis to assess their impact on the overall findings.

2.5. Selection of studies using preferred reporting items for systematic reviews and meta-analyses flowchart (PRISMA)

The reporting items for systemic literature review usually identified the PRISMA flowchart, as it provides evidence-based items for results synthesis. The PRISMA flowchart highlights the number of relevant literature reviews eligible to provide significant information related to



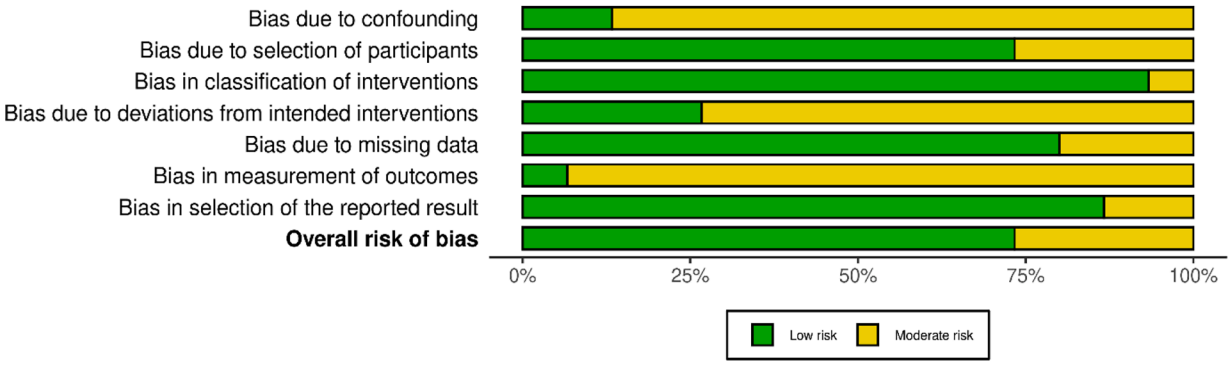


Fig. 2. Risk of bias assessment by ROBINS-1 tool.

the research question. The flowchart, Fig. 1 also includes the excluded articles and duplicate literature reviews.

2.6. Study selection

The initial search yielded 297 articles. After removing duplicates, 297 articles remained. Titles and abstracts were screened by two independent reviewers to identify potentially relevant studies. Disagreements were resolved through discussion or consultation with a third reviewer. After the initial screening, 97 articles were selected for full-text review.

During the full-text review, 62 articles were excluded based on the predefined criteria, resulting in 15 studies being included in the final qualitative synthesis.

2.7. Data extraction

Data were extracted from the included studies using a standardized extraction form. The form captured information on:

- Study characteristics (authors, publication year, and journal).
- AI technology and application in healthcare.
- Aspects of AI safety, trust, security, or responsible use addressed.
- Key findings and conclusions.

Two reviewers independently extracted data to ensure accuracy and consistency. Any discrepancies were resolved through discussion or by consulting a third reviewer.

2.8. Data synthesis

Data synthesis involved a narrative summary of the findings from the included studies, organized around the key themes of AI safety, trust, security, and responsible use. The synthesis aimed to:

- Summarize the current state of research.
- Identify gaps in the literature.
- Highlight emerging trends and technologies.
- Provide recommendations for future research.

2.9. Risk-of-bias assessment

Risk-of-bias assessments have been conducted using ROBINS-I (for OS) and it has been projected as a traffic-light plot using the online platform robvis™ to create figures for the quality of risk-of-bias assessment, shown in Fig. 2.

3. Results

The reviewed studies offer a comprehensive look at the integration of Artificial Intelligence (AI) in healthcare, covering a range of issues from trust and safety to ethics and practical implementation. The outcomes can be prearranged chronologically, reflecting the evolving understanding of AI's role in this critical field.

3.1. Early investigations (2000–2020)

The foundation for understanding trust in AI was laid by studies like those of Lockey et al. [5], who recognized several key challenges distinctive to AI systems. Their work underlined the significance of transparency, explainability, accuracy, and the balance amongst automation and human augmentation. They correspondingly jagged out the vulnerabilities tackled by different stakeholders, emphasizing the need for a multi-stakeholder approach to constructing trustworthy AI systems. On the other hand, the real-world employment of such a methodology remains a significant challenge, for the most part when balancing transparency with the need to protect data privacy and proprietary algorithms.

3.2. Mid-phase studies (2017–2021)

As the field evolved, the attention shifted in the direction of more practical aspects of AI deployment. For instance, Cutillo et al. [6] stressed the importance of explainability and usability in AI systems, arguing that these factors are fundamental for fostering trust among healthcare providers. They underlined the need for AI systems to be not only accurate but also interpretable and user-friendly. Conversely, they also accredited the potential trade-offs between explainability and predictive accuracy, especially in complex AI models.

Shneiderman's [7] conceptual work delivered a governance framework for Human-Centered AI (HCAI), offering practical steps to upsurge the reliability, safety, and trustworthiness of AI systems. His commendations, while valuable, were noted for deficient experimental validation, and there was concern that overly prescriptive governance could suppress innovation in AI development.

Ellahham et al. [8] engrossed on the safety risks associated with AI, such as distributional shifts and data quality issues. They advocated for safety strategies like "safe-fail" designs to mitigate these risks. However, their work was primarily theoretical, pointing to the need for empirical studies to validate these safety measures in real-world settings.

Esmaeilzadeh's [9] study offers important understandings into public perceptions of AI in healthcare, emphasizing how concerns about technology, ethics, and regulation impact the willingness to adopt AI-based tools. Key findings disclose that technological anxieties, such as potential errors and lack of transparency combined with ethical uncertainty and regulatory reservations, significantly impact public acceptance.

3.3. Recent insights (2021–2022)

More recent studies have explored into the practical application of AI in healthcare and its ethical implications. Asan et al. [10] explored clinician trust in AI, suggesting that trust is influenced by factors like data quality, system transparency, and the reliability of AI outputs. They introduced the concept of "optimal trust," where clinicians maintain a balanced skepticism toward AI recommendations. However, implementing this model in practice, especially in diverse clinical environments, remains a challenge.

Choudhury et al. [11] conducted a systematic review that demonstrated AI's potential to improve patient safety outcomes, particularly in clinical alarms and drug safety. Nonetheless, they also identified significant challenges, such as the lack of standardization in AI performance reporting, which could hinder the consistent application of AI across different healthcare settings.

Nagendran et al. [12] provided a critical evaluation of AI performance in diagnostic imaging, revealing that many studies overstated AI's effectiveness compared to human clinicians. Their findings highlighted the need for more rigorous methodologies and better reporting standards to ensure that AI's capabilities are accurately represented.

Rodriguez et al. [13] proposed a comprehensive framework for trustworthy AI, integrating ethical principles with regulatory and technical requirements. While their framework is theoretically sound, its practical application across diverse cultural and regulatory contexts remains untested.

In the realm of patient interactions, Milne-Ives et al. [14] found that conversational agents in healthcare were generally well-received, with high levels of user satisfaction. However, they noted limitations in these agents' language understanding and interactivity, which could impact their effectiveness and patient trust.

Richardson et al. [15] added another dimension by exploring patient perspectives on AI. Their findings revealed significant concerns about safety, data integrity, and potential cost implications of AI in healthcare.

This study stressed the importance of addressing patient concerns proactively to form public trust in AI technologies.

Nadella et al. [16] underscored the progressions in AI-driven diagnostic tools and predictive analytics, despite the fact that correspondingly indicating the gaps in regulatory frameworks and the need for unbiased AI models. Their review put forward that, despite the progress, there is still much work to be done to ensure the ethical and effective integration of AI in healthcare.

Another study by Naik et al. [17], aimed to discuss and define the place of AI in healthcare, with a specific focus on some of the legal and ethical considerations arising. The authors found that AI is very promising in several applications, such as diagnostics and drug discovery, but its use raises many accountability concerns, such as privacy and data protection. Probably the greatest single challenge facing the clinical adoption of AI out of all the challenges is the lack of standardized regulations, particularly with regard to algorithmic transparency and bias. This review has called for instating a legal framework that protects patients and ensures the AI systems are ethical and remain within the legal ambit.

Markus et al. [18], provided a general overview of the development of explainable AI systems and their potential to improve trust in healthcare applications of AI. The authors emphasized the balance between interpretability and model performance, plus the need for transparency to foster trust among healthcare providers. They discussed how explainable AI would be key to further improving clinician confidence, especially around decision-making. Although explainability looks very promising, it does need further empirical support to prove the practical benefits. The study also recommended complementary measures:

Table 1
Summary of the studies included.

Study no.	Year	Focus	Key findings	Challenges identified	Contributions
01. Lockey et al. [5]	2000–2020	Trust in AI Systems	Identified key trust challenges including transparency, explainability, accuracy, and automation vs. augmentation.	Practical implementation of multi-stakeholder approach; Balancing transparency with privacy	Established foundational understanding of trust challenges in AI; Advocated for multi-stakeholder approach
02. Cutillo et al. [6]	2017	Explainability and Usability	Emphasized the importance of explainability and usability for fostering trust among healthcare providers.	Potential trade-offs between explainability and predictive performance	Highlighted critical role of user-centered design in AI; Contributed to explainable AI design principles
03. Shneiderman [7]	2019	Governance of Human-Centered AI (HCAI)	Proposed a three-layer governance structure to improve AI reliability, safety, and trustworthiness.	Lack of empirical validation; Potential risk of overly prescriptive governance	Provided a comprehensive governance framework for HCAI; Introduced practical steps for enhancing AI safety
04. Ellahham et al. [8]	2020	Safety Concerns in AI Deployment	Identified safety risks like distributional shifts and data quality issues; Advocated for "safe-fail" designs.	Need for empirical validation of safety strategies; Potential slowing of AI adoption	Provided a thorough review of AI safety challenges; Proposed safety strategies for AI deployment
05. Esmaeilzadeh [9]	2020	Perceptions of AI in Healthcare	Explored how technological, ethical, and regulatory concerns affect the perceived risks of AI in healthcare.	Anxiety over AI performance and ethical implications; Concerns about regulatory gaps	Developed a model to understand factors influencing perceived risks and benefits in AI-based healthcare tools.
06. Asan et al. [10]	2021	Clinician Trust in AI Systems	Explored clinician trust influenced by data quality, transparency, and reliability; Introduced "optimal trust" model.	Implementing "optimal trust" across varied clinical environments; Balancing skepticism and reliance	Developed the concept of "optimal trust" in AI; Enhanced understanding of factors influencing clinician trust
07. Choudhury et al. [11]	2021	AI in Patient Safety Outcomes	Demonstrated AI's potential to improve patient safety, particularly in clinical alarms and drug safety.	Lack of standardization; Variability in AI performance reporting	Systematic review of AI's impact on patient safety; Highlighted both benefits and limitations in AI applications
08. Nagendran et al. [12]	2021	AI Performance in Diagnostic Imaging	Found overstated AI performance in diagnostic imaging; High risk of bias in non-randomized trials.	Need for more rigorous methodologies and better reporting standards	Provided critical evaluation of AI performance in imaging; Called for more robust and unbiased research
09. Rodriguez et al. [13]	2021	Holistic Framework for Trustworthy AI	Proposed an integrated framework combining ethics, regulation, and technical requirements for AI trustworthiness.	Lack of empirical data; Challenges in applying the framework across diverse contexts	Developed a holistic framework for AI trustworthiness; Aimed to align AI development with societal values
10. Milne-ives et al. [14]	2021	Conversational Agents in Healthcare	Positive user reception; High usability and satisfaction, but limitations in language understanding and interactivity were noted.	Limitations in language processing and interactivity could hinder trust and effectiveness	Reviewed effectiveness of conversational agents; Identified areas for improvement in user interaction and satisfaction
11. Richardson et al. [15]	2022	Patient Perspectives on AI	Patients expressed concerns about safety, data integrity, and potential cost implications of AI.	Limited generalizability due to demographic focus; Need for more diverse patient perspectives	Provided insights into patient concerns and expectations; Highlighted the importance of addressing patient-centered issues in AI
12. Nadella et al. [16]	2022	AI and ML Applications in Healthcare	Highlighted advancements in diagnostic tools and predictive analytics; Identified gaps in regulatory frameworks and the need for unbiased models.	Overlooked long-term AI trends; Challenges in integrating AI with existing healthcare systems	Comprehensive review of AI/ML applications; Emphasized the need for ethical and regulatory improvements in AI integration
13. Naik et al. [17]	2022	Legal and Ethical Implications of AI	Highlighted ethical challenges such as informed consent, transparency, and cybersecurity; Called for comprehensive governance frameworks.	Theoretical focus limits practical application; Implementing governance across diverse settings	Comprehensive review of legal and ethical issues; Framework for governance in AI healthcare
14. Markus et al. [18]	2021	Explainability in AI	Proposed a framework for selecting explainability methods based on healthcare context.	Trade-offs between explainability and predictive accuracy; Complexity in method selection	Contributed to the formalization of explainable AI (XAI); Provided guidance for choosing appropriate explainability methods
15. Habli et al. [19]	2021	Moral Accountability and Safety	Analyzed how AI-based tools challenge traditional models of clinical decision-making and accountability.	Weakened traditional accountability models in clinical decisions involving AI	Called for updated frameworks to address moral accountability and safety in AI-driven clinical decision-making.

external validation and standardized regulations to guarantee more trustworthiness in AI models.

Habli et al. [19], reviewed some ethical and safety concerns regarding the use of AI in clinical decision-making for sepsis treatment. The AI Clinician system developed at Imperial College London demonstrated an opportunity for the personalization of fluid and vasopressor management in patients with sepsis. However, this study identified considerable deficits concerning moral accountability since clinicians do not know or control AI-driven decisions. The authors present the case for a dynamic model of safety assurance that embeds ongoing monitoring of AI systems post-deployment and calls for developers to be included in accountability frameworks to offer added protection to patients.

4. Summary

The progression of AI in healthcare research discloses a clear evolution from identifying theoretical challenges to addressing practical implementation issues. Trust, safety, and ethics have appeared as central themes, with each study contributing to a deeper understanding of these complex issues. Conversely, the results also point out that significant challenges remain, mainly in the areas of empirical validation, regulatory oversight, and the balance between innovation and ethical practice. As AI continues to grow, addressing these challenges will be decisive for comprehending its full potential in improving healthcare outcomes. Table 1, presents a summary of fifteen studies included in the study.

5. Discussion

The incorporation of AI in healthcare is a swiftly developing field with noteworthy prospective to transform medical practices, patient care, and healthcare systems. On the other hand, this potential approaches with numerous challenges, predominantly regarding trust, safety, ethical considerations, and the practical implementation of AI technologies. This discussion critically investigates the outcomes from all the studies included, exploring common themes, divergences, and the broader implications for AI in healthcare. By scrutinizing these studies communally, this section aims to provide a wide-ranging understanding of the current state of AI in healthcare and to recognize areas where further research and development are needed.

5.1. Trust and acceptance of AI in healthcare

Trust is a central theme across many of the studies, replicating its significance in the effective adoption and employment of AI technologies in healthcare. Asan et al. [10] provided a thorough investigation of trust between clinicians and AI systems, laying emphasis on the need for an “optimal trust” model. This model proposes that clinicians should maintain a balanced level of skepticism towards AI outputs, which is fundamental for avoiding over-reliance on AI. Nevertheless, this approach presents practical challenges, predominantly in safeguarding that clinicians have the necessary understanding of AI’s capabilities and limitations. The study’s focus on trust is significant, but it could be criticized for not sufficiently addressing how trust can be cultivated in environments where clinicians may have variable levels of understanding with AI technologies.

Cutillo et al. [6] also emphasized trust, particularly focusing on the significance of explainability, usability, and transparency in AI systems. The authors say that these factors are important for building trust among healthcare providers. Despite the fact that this standpoint is critical, it undertakes that AI developers will highlight user-centered design, which may perhaps not all the time align with commercial or technological priorities. Likewise, the study’s dependence on explainability as a key to trust could be seen as more than usually optimistic, particularly in cases wherever full transparency may not be conceivable due to the complex nature of AI algorithms.

Lockey et al. [5] extended the discussion by recognizing five key trust challenges specific to AI: transparency, explainability, accuracy, automation versus augmentation, and mass data extraction. Their suggestion for a multi-stakeholder approach to building trustworthy AI is comprehensive but raises questions about the feasibility of such a method in practice. The contribution of various stakeholders, each with diverse priorities and concerns, possibly will confound efforts to form a unified framework for trust. Likewise, the task of maintaining balance in transparency with data privacy and proprietary concerns remains a noteworthy hurdle that needs more concrete solutions.

5.2. Safety and ethical concerns

The safety of AI systems is a foremost concern, as highlighted by Ellahham et al. [8] and Habli et al. [19] The study by Ellahham et al. [8], discussed the possible safety risks linked with AI in healthcare, such as distributional shifts, data quality issues, and the unpredictability of AI behavior. Their advocacy for safety strategies like “safe-fail” designs is critical but then again might not fully address the complexities of real-world AI deployment. The stress on theoretical safety measures deprived of empirical validation leaves a gap in understanding how these strategies will perform in practice. Furthermore, the study could be appraised for not addressing the possible trade-offs between safety and innovation, where excessively watchful approaches might suppress the development of beneficial AI applications.

Habli et al. [19] study suggested a nuanced analysis of moral accountability in AI-driven clinical decision-making. The study challenges traditional ideas of responsibility, arguing that AI’s involvement in clinical decisions abates the conditions for moral accountability. This is a significant influence to the discourse on AI ethics, as it highlights the necessity for new frameworks that contemplate the roles of AI developers and safety engineers. Still, the study could benefit from more concrete proposals on how these new models of accountability might be structured and enforced, primarily in complex healthcare environments where multiple actors contribute to decision-making processes.

Ethical concerns are further explored by Naik et al. [17] and Rodriguez et al. [13] The study by Naik et al. [17] focuses on the legal and ethical challenges of AI in healthcare, encouraging for comprehensive governance frameworks that address issues such as informed consent, algorithmic transparency, and cybersecurity risks. Although their recommendations are timely and relevant, the study’s theoretical nature limits its applicability to various healthcare settings where supervisory environments and resources vary widely. This could result in an overly rigid approach to regulatory frameworks, less supple, and hence less able to bend with the incredibly fast pace of AI innovation.

Under Rodriguez et al. [13], a framework for trustworthy AI is fully put forward that effectively links ethical principles with regulatory and technical requirements. This is very laudable from the point of view of comprehensiveness, but there are also practical questions about how this can be implemented. The focus on the alignment of AI development with societal values is very key in this study, but operationalizing these principles within differing cultural and regulatory contexts may not be plain sailing. Also, the shortage of experimental data in this study marks it difficult to get an understanding of how effectively this framework would work in real world settings.

5.3. Patient perspectives and public trust

Understanding patient standpoints is essential for the effective incorporation of AI in healthcare. Richardson et al. [15] explored patient attitudes toward AI, illuminating concerns about safety, data integrity, and the potential impact on healthcare costs. The study’s findings put the spotlight on the need to take into account the concerns expressed by patients to gain public trust in AI technologies. However, when generalizing from this study, one of the major concerns is it was focused on only one demographic group—mostly white, non-Hispanic participants. This constraint put forward a need for more diverse and representative

studies to capture the full spectrum of patient perspectives on AI in healthcare.

Milne-Ives et al. [14] provide a generally positive assessment of conversational agents in healthcare, reporting high usability and satisfaction among users. However, the study also notes limitations, such as poor language understanding and limited interactivity, which could affect patient trust and satisfaction. The mixed effectiveness of these agents highlights the need for more rigorous evaluations of AI technologies, particularly in understanding the factors that contribute to or detract from patient trust. In addition, the study's focus on conversational agents may manage other AI applications in healthcare that may perhaps present different challenges and opportunities for patient engagement.

Additional study by Esmaeilzadeh [9], which surveyed 307 individuals in the United States, offers valuable understandings into public perceptions of AI in healthcare. The research detects technological concerns, such as performance and communication features as the most substantial prognosticators of perceived risk. Ethical issues, including trust factors and transparency, and regulatory concerns regarding data privacy and accountability also considerably impact public attitudes. The study discloses that technological, ethical, and regulatory concerns collectively shape public perceptions of AI in healthcare. The outcomes suggest that addressing these concerns through enhanced transparency, obedience to ethical standards, and the development of robust regulatory frameworks is critical for growing public trust and acceptance. The study's limitations include potential bias from confounding variables and the reliance on self-reported data, which may not completely capture various public opinions. The sample may also not be representative of the broader population, which could affect the generalizability of the outcomes. To increase public understanding and recognition of AI in healthcare, future research must focus on a more diverse and representative sample of participants. Furthermore, exploring how different demographic groups perceive AI and evaluating interventions to enhance public education about AI technologies are important steps. Addressing these areas can aid in improving regulatory guidelines and ethical standards to better align with public concerns.

5.4. Advancements and limitations of AI in healthcare

The potential of AI for successful diagnostic accuracy and patient outcomes is well documented in publications such as Choudhury et al. [11] and Nagendran et al. [12]. In the study of Choudhury et al. [11], they discussed the demonstration of the application of A.I, as it has immense potential for considerably improving patient safety outcomes linked to clinical alarms and drug safety. They also point out that major challenges remain to be overcome, such as a lack of standardization and variability in the reporting of AI performance, which might undermine consistency in care. Critical here is that while AI might improve health care, its benefits may not be felt uniformly across all settings and patient populations.

The paper by Nagendran et al. [12] criticizes standards in deep learning research, particularly the high risk of bias inherent in non-randomized studies. Their findings implicate overstatement of AI's performance relative to clinicians that inflames the hype about AI in health care. This critique is important mainly because it requires evidence-based approaches toward the evaluation of AI technologies and ensuring that robust data support such claims of AI superiority. The study can, however, be criticized in that it focused on deep learning algorithms for diagnostics, just one piece of what AI can do in health.

Nadella et al. [16] further contribute to this update regarding AI developments with emphasis on ethical and regulatory gaps that should be filled for the responsible incorporation of AI. The review identifies, on the one hand, some of the really transformative potential of AI, and it also brings out significant concerns in privacy, algorithm transparency, and the need for a strong governance framework. Though their recommendations are complete, the focus of this study on recent literature may miss out on the longer-term trends and development regarding AI, which would be helpful in identifying how these challenges have evolved.

[20]. The exposed files included very sensitive information, such as medical records, passports, and national identification information. The consequences of this security breach went beyond simple privacy issues; it presented considerable threats related to identity theft, fraudulent activities, and the possible exploitation of medical data, which forced reevaluation of the effectiveness of existing cybersecurity measures in protecting healthcare artificial intelligence systems.

This case highlights the bifacial nature of artificial intelligence in the health sector. While artificial intelligence brings unparalleled levels of efficiency and accuracy to the administration of patient care, it also requires commensurately sophisticated cybersecurity measures to protect the data on which it depends. One of the most important lessons learned from the WotNot breach is that storage and management practices for data within healthcare solutions using artificial intelligence must be strictly monitored. Misconfigurations, as seemingly minor as an improperly secured cloud storage bucket, can lead to catastrophic consequences, undermining the trust of patients and providers alike.

5.5. The role of explainability and usability

This is again an important point associated with the explanatory power and usability of AI systems, more specifically elaborated in Markus et al. [18], who review the role of explainability in creating trustworthy AI for healthcare. This would provide an overview of different XAI methods and offer guidance on how to choose from different explainability techniques in given healthcare contexts. Such contribution is important in handling the challenge of developing AI systems that are not only accurate but also interpretable for both clinicians and patients. However, the paper gives focal attention to explainability, which understates potential trade-offs between explainability and predictive performance. This may turn out to be particularly true with complex AI systems where increasingly transparent models are known to lessen model accuracy.

Shneiderman [7] also links the problem of explainability to a larger framework of human-centered AI. The study provides actionable recommendations to increase the reliability, safety, and trustworthiness of HCAI systems regarding audit trails, verification and validation testing, and bias testing. Although this framework by Shneiderman [7] does give a very useful sense of some of the specific, actionable steps that can be taken toward creating better AI systems, a clear limitation is that these recommendations are not based on empirical evidence.

5.6. The future of A.I in healthcare: striking a balance between innovation and regulation

Nadella et al. [16] and Nagendran et al. [12] pointed out the challenge of striking a balance between these two approaches to developing AI: a process that would underpin innovation with ethical and regulatory oversight. Nadella et al. [16] paid great attention to developing unbiased AI models and achieving interoperability with existing healthcare systems. Nagendran et al. [12] underscore the increased stringency of reporting standards and transparency in AI research. All of these studies together suggest that the pathway to AI will have to be a shared one, bringing technologists, healthcare providers, regulators, and patients all into the same space. The challenge is, therefore, between fostering innovation and ensuring that AI systems are safe, ethical, and effective. A balance in these will be very key in setting a future pathway for AI in health.

6. Future recommendations

1. Ensuring trust and security in AI applications in healthcare

Major opportunities for improving patient outcomes and operational

efficacies has come into being by the introduction of AI into healthcare. On the other hand, the successful acceptance of AI applications centers on establishing trust and security among healthcare providers, patients, and stakeholders. This discussion makes results from various studies to shape best practices for designing AI applications that foster trust and ensure security in healthcare settings.

2. Understanding trust in AI systems

Trust in AI systems is a complex construct that would greatly influence acceptance and the degree of adoption of such technologies in health. According to Asan et al. [10], there are three primary dimensions to form trust: benevolence, integrity, and ability. These dimensions concern clinicians' perceptions of AI systems and their capabilities. A proper understanding of these factors is important for the design of AI applications that clinicians can rely on for clinical decision-making.

AI systems should support transparency and explainability. Cutillo et al. [6] underline the need for explainability, intended as interpretability. It is clear that insights from AI decision-making processes can be obtained by developers and passed on to clinicians and patients in order to engender trust in the rationale behind AI recommendations.

Furthermore, the "optimal trust" outlined by Asan et al. [10] is that both the clinicians and AI systems have to be equally skeptical about the outputs provided by one another. With such a sensible approach there will be no over dependence on AI recommendations. As a result, the clinician stays involved in the process of decision making with a critical evaluation of the AI suggestions.

3. Ensuring security and safety

The safety and security of AI applications in healthcare are vital. Ellahham et al. [8] identify several safety concerns associated with AI deployment, including distributional shifts, poor data quality, and uncertainty in predictions. AI systems must be designed with strong safety protocols and procedural safeguards to decrease these risks.

As per the incident of WotNot ChatBot [20] incident, these would need to be addressed in a multi-faceted approach. First, healthcare organizations need to institute anticipatory security measures: end-to-end encryption, role-based access control, and routine audits of data systems. While these practices are normative in cybersecurity, they need to be attuned to the specific demands of healthcare, where data integrity and confidentiality have a direct bearing on patient outcomes.

Second, developers of healthcare AI systems should incorporate security features at the design phase so that safeguards are not just an afterthought but inherent to the architecture of the system, including robust failsafes that can detect and neutralize threats before sensitive data is compromised.

Lastly, the WotNot [20] breach highlights the clear demands for transparency and accountability within healthcare. Patients and professionals in healthcare need to have confidence that not only will the AI systems work but that there is also commitment to keeping the data safe. Only through proper communication of how data is stored, processed, and secured can this trust be restored, in the wake of breaches such as this one.

Furthermore, regulatory bodies have to enforce more stringent compliance standards for healthcare AI systems so that they are resilient against the evolving cyber threats. The breach involving WotNot [20] is a cautionary tale put into practice, highlighting new vulnerabilities that arise with the rapid integration of artificial intelligence in the healthcare sector. While the technology holds many promises, its effectiveness can only be realized with sufficient trust and security. All these events are lessons in themselves, and by adoption of better cybersecurity measures, healthcare organizations can really benefit from the promises of AI with privacy and welfare protection extended to patients.

6.1. Key strategies for ensuring safety include

1. **Safe Design Principles:** AI applications should make safety an integral part of their development. In-built safety margins should have fail-safe mechanisms to revert the system to human control in case of any unexpected behavior or errors.
2. **Rigorous Testing and Validation:** The testing required to ensure the performance of AI systems in highly variable clinical scenarios should be rigorous before their deployment. This would involve validation of AI algorithms against real-world data for their reliability and accuracy.
3. **Continuous Monitoring and Evaluation:** There will be a requirement for continuous assessment of the AI system in the course of clinical practice to recognize any safety issues that may ascend and get used to changes in the clinical environment. This is dynamism toward safety assurance, recognizing that risks may evolve over time.

6.2. Addressing ethical considerations

Furthermost of the ethical concerns in AI in health—bias, privacy, accountability—must be taken into deliberation in order to form trust and safety. Choudhury et al.[11] raise the question of how AI algorithms should be fair to return unbiased results for different types of patients. Developers should emphasize making use of representative datasets and design bias mitigation strategies to deter discrimination driven by AI outputs.

Moreover, AI decision-making drives important ethical implications and questions around moral accountability. Traditionally, according to Habli et al.[19], accountability may not apply in the context of AI since clinicians are usually very limited in control over AI-made decisions. In that respect, developers of AI and safety engineers have to be involved in the discussion on accountability for patient harm done by AI systems.

6.3. Enhancing the education and training of users

This is the point where education and training of users become vital for gaining trust as well as ensuring responsible use. Asan et al.[10] underline that large differences in past experiences and familiarity with AI systems are critical reasons for the differences in levels of trust by clinicians. In view of that, the emphasis of the training programs needs to be on improving the understanding regarding the capabilities, limitations, and ethical considerations of AI by clinicians.

6.4. These training initiatives can include the following

1. Workshops and Seminars: Getting the healthcare professional involved in discourses on AI technologies, their application, and associated ethics can help build confidence in working with these systems.
2. Simulation-Based Training: The experience of an AI tool in a simulated environment could help clinicians develop comfort working with AI applications and further their understanding of how the applications work.
3. Disciplinary Collaboration: Aiding to assist and collaboration among developers of AI systems, clinicians and ethicists. This will allow the development of AI systems to meet not only both clinical needs but also high standards of ethics.

6.5. Regulatory frameworks and standards

It is unavoidable upon clear regulatory frameworks to ensure that AI is established and deployed in health care with safety and ethical considerations. Guidelines will have to be adaptive as evolution occurs in technologies or when challenges first appear. This would call for stakeholders to stand up collected and frame guidelines at a time when

patient safety and ethical considerations in evolving and deploying AI take up a front-row seat.

6.6. Key considerations for regulatory frameworks include

1. Standardization of Performance Metrics:

There should be standard levels to calculate the performance of AI, which stretches credibility to the comparison between studies and applications, as a consequence making them transparent and accountable.

2. Clear-cut Guidance on Deployment of AI:

The regulatory bodies must bring forward clear guidelines for the safe utilization of the AI system with protocols for testing, validation, and continuing surveillance.

3. Involvement of Diverse Stakeholders:

Engaging a wide variety of stakeholders, from clinicians and patients to AI developers, in the regulatory process could help guarantee that diverse viewpoints are taken into deliberation while writing standards.

7. Conclusion

The studies appraised contribute a comprehensive and sophisticated understanding of the potential and challenges for AI in healthcare. The most prominent themes that emerge are those of trust, safety, ethics, and explainability, each highlighting very complex problems. From this analysis, several gaps and limitations have been noticed, mainly in terms of the practicality of application for the proposed frameworks and strategies. In this respect, empirical validation, the challenge of balancing transparency with proprietary concerns, and the complexities around patient trust all present avenues for further research.

AI holds great promise for improving healthcare outcomes, optimizing operational efficiency, and advancing personalized medicine. However, the integration of AI into healthcare systems is also burdened with immense challenges in transparency, trust, security, and ethical considerations. The findings of this study underscore the need for a holistic approach to ensure that AI systems used in health are not only effective but also have a characteristic of equity, safety, and reliability.

In general, it is about the responsible adoption of AI technologies that involves favoring those with a strong transparency feature, including XAI, and the adequate training of healthcare providers in understanding the capabilities and limitations of AI tools. Also, the integration of AI into clinical workflows has to be designed in a way to highlight human oversight that will minimize risks and cultivate confidence from both clinicians and patients.

Policymakers should focus on developing agile regulatory frameworks that address modern issues of data privacy, algorithmic bias, and cybersecurity threats. Regulations need to mandate extensive testing of AI systems before their deployment and define clear standards for accountability in case of errors or security breaches. Furthermore, federated learning and similar developments offer practical solutions to protect privacy and should be incentivized through supportive policy measures.

There must be empirical assessments of the AI systems in authentic healthcare settings at the heart of most research going forward. Also, key further areas are looking into federated learning, first across multi-institutional frameworks and then defining and assessing the fairness evaluation metrics for AI, to target healthcare-specific adversarial vulnerabilities. Further, given the broad impact of health-relevant AI, addressing this change requires interdisciplinary engagement with ethicists, legal scholars, social scientists, and those skilled in governance.

By addressing such actionable priorities, stakeholders can collaboratively create an environment where AI technologies can thrive in a secure, ethical, and socially responsible manner. This holistic approach will ensure that the potential of AI in healthcare is fully realized while safeguarding the trust and well-being of both patients and providers.

8. Ethics and Consent

Ethical approval and consent were not required.

9. Underlying Data

No Data are associated with this article.

10. Reporting Guidelines

PRISMA 2020 checklist.

11. Grant information

The author(s) declared that no grants were involved in supporting this work.

Author contributions

- **Muhammad Mohsin Khan** led the conceptualization and design of the study. He oversaw the project's methodology, coordinated team efforts, and ensured the review followed strict scientific standards from start to finish.
- **Noman Shah** developed and implemented the search strategy for the literature review. He handled the screening process, selected studies based on eligibility criteria, and performed quality assessments to make sure only the most relevant studies were included.
- **Nissar Shaikh** was in charge of data extraction and quality control. He carefully pulled key data from the selected studies, checked for accuracy, and organized everything so that the analysis could proceed smoothly.
- **Abdulnasser Thabet** took the lead on analyzing and synthesizing the data. He reviewed the extracted data, identified trends, and created tables to clearly present the main findings of the review.
- **Talal Alrabayah** wrote the first draft of the manuscript, crafting a cohesive narrative that brought together the findings and conclusions of the review in a clear and structured format.
- **Sirajeddin Belkhair** reviewed and edited the draft, providing critical feedback to improve clarity and coherence. He also facilitated access to resources needed to complete the review.

Each author played a key role in bringing this work to completion, and they all reviewed and approved the final version of the manuscript. The authors take collective responsibility for the accuracy and integrity of the work presented.

Ethical Approval

This study was a systematic review of previously published literature, so no ethical approval or informed consent was necessary.

CRediT authorship contribution statement

Muhammad Mohsin Khan: Methodology, Conceptualization. **Noman Shah:** Writing – original draft, Resources, Methodology. **Nissar Shaikh:** Formal analysis, Data curation. **Abdulnasser Thabet:** Validation, Formal analysis. **Talal alrabayah:** Writing – original draft. **Sirajeddin Belkhair:** Writing – review & editing, Validation, Supervision, Project administration.

Funding

This research was conducted without any funding support.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] F. Li, N. Ruijs, Y. Lu, AI: A Systematic Review on Ethical Concerns and Related Strategies for Designing with AI in Healthcare, 2023.
[2] HITRUST Alliance. Navigating the Security Risks of AI in Healthcare. Retrieved from https://hitrustalliance.net/blog/navigating-the-security-risks-of-ai-in-healthcare.
[3] G. Karimian, E. Petelos, S.M.A.A. Evers, The ethical issues of the application of artificial intelligence in healthcare: a systematic scoping review, 2022.
[4] Forbes. Ethical AI in Healthcare: A Focus on Responsibility, Trust, and Safety. Retrieved from https://www.forbes.com/sites/forbesbooksauthors/2024/01/04/ethical-ai-in-healthcare-a-focus-on-responsibility-trust-and-safety/.
[5] S. Lockey, N. Gillespie, D. Holm, I.A. Someh, A review of trust in artificial intelligence: Challenges, vulnerabilities, and future directions. Proceedings of the Hawaii International Conference on System Sciences (HICSS), Honolulu, HI, United States, 4-8 January 2021. doi: 10.24251/hicss.2021.664.
[6] C.M. Cutillo, K.R. , Sharma, L. Foschini, S. Kundu, M. Mackintosh, K.D. Mandl, MI in Healthcare Workshop Working Group. Machine intelligence in healthcare—perspectives on trustworthiness, explainability, usability, and transparency, npj Digital Med., 3(1) (2020) 5. Doi: 10.1038/s41746-020-0254-2.
[7] Ben Shneiderman, Bridging the Gap between Ethics and Practice: Guidelines for Reliable, Safe, and Trustworthy Human-centered AI Systems. ACM Trans. Interact. Intell. Syst. 10(4) (2020) Article 26 (October 2020), 31 pages. Doi: 10.1145/3419764.
[8] S. Ellahham, N. Ellahham, M.C.E. Simsekler, Application of artificial intelligence in the health care safety context: opportunities and challenges, Am. J. Med. Qual. 35 (6) (2020) 341–348, https://doi.org/10.1177/1062860620907235.
[9] P. Esmaeilzadeh, Use of AI-based tools for healthcare purposes: a survey study from consumers' perspectives, BMC Med. Inf. Decis. Making 20 (1) (2020) 1–19, https://doi.org/10.1186/s12911-020-01191-1.
[10] O. Asan, A.E. Bayrak, Choudhury A. Artificial Intelligence and Human Trust in Healthcare: focus on Clinicians J. Med. Internet Res. 22(6) (2020) e15154 URL: http://www.jmir.org/2020/6/e15154/ doi: 10.2196/15154 PMID: 32558657.
[11] A. Choudhury, O. Asan, Role of Artificial Intelligence in Patient Safety Outcomes: Systematic Literature Review JMIR Med Inform. 8(7) (2020) e18599, URL: http://medinform.jmir.org/2020/7/e18599/ doi: 10.2196/18599 PMID: 32706688.
[12] M. Nagendran, Y. Chen, C.A. Lovejoy, A.C. Gordon, M. Komorowski, H. Harvey, E. J. Topol, J.P.A. Ioannidis, G.S. Collins, M. Maruthappu, Artificial intelligence versus clinicians: systematic review of design, reporting standards, and claims of deep learning studies, *BMJ* 368 (2020) m689, https://doi.org/10.1136/bmj.m689.
[13] N. Díaz-Rodríguez, J. Del Ser, M. Coeckelbergh, M. López de Prado, E. Herrera-Viedma, F. Herrera, Connecting the dots in trustworthy Artificial Intelligence: from AI principles, ethics, and key requirements to responsible AI systems and regulation, *Inf. Fusion* 99 (2023) 101896, https://doi.org/10.1016/j.inffus.2023.101896.
[14] M. Milne-Ives, C. de Cock, E. Lim, M.H. Shehadeh, N. de Pennington, G. Mole, E. Normando, E. Meinert, The Effectiveness of Artificial Intelligence Conversational Agents in Health Care: Systematic Review J Med Internet Res 2020;22(10):e20346 URL: http://www.jmir.org/2020/10/e20346/ doi: 10.2196/20346 PMID: 33090118.
[15] J.P. Richardson, C. Smith, S. Curtis, S. Watson, X. Zhu, B. Barry, R.R. Sharp, Patient apprehensions about the use of artificial intelligence in healthcare, *npj Digital Med.* 4 (1) (2021) 140, https://doi.org/10.1038/s41746-021-00509-1.
[16] G.S. Nadella, S. Satish, K. Meduri, S.S. Meduri, A systematic literature review of advancements, challenges and future directions of AI and ML in healthcare, Int. J. Mach. Learn. Sustain. Develop. 5 (3) (2023) 123–145, https://doi.org/10.1016/j.ijmlsd.2023.100123.
[17] N. Naik, B.M.Z. Hameed, D.K. Shetty, D. Swain, M. Shah, R. Paul, K. Aggarwal, S. Ibrahim, V. Patil, K. Smriti, S. Shetty, B.P. Rai, P. Chlosta, B.K. Somani, Legal and ethical consideration in artificial intelligence in healthcare: Who takes responsibility? *Front. Surg.* 9 (2022) 862322 https://doi.org/10.3389/fsurg.2022.862322.
[18] A.F. Markus, J.A. Kors, P.R. Rijnbeek, The role of explainability in creating trustworthy artificial intelligence for health care: a comprehensive survey of the terminology, design choices, and evaluation strategies, *J. Biomed. Inform.* 108 (2020) 103655, https://doi.org/10.1016/j.jbi.2020.103655.
[19] I. Habli, T. Lawton, Z. Porter, Artificial intelligence in health care: accountability and safety, *Bull. World Health Organ.* 98 (4) (2020) 251–256, https://doi.org/10.2471/BLT.20.253135.
[20] CyberNews. (2024, February 15). WotNot exposes 346K sensitive customer files. Retrieved from https://cybernews.com/security/wotnot-exposes-346k-sensitive-customer-files/.