

Proceeding Paper

A Comprehensive Framework for Transparent and Explainable AI Sensors in Healthcare [†]

Rabaï Boudershem *

College of Law, Prince Mohammad Bin Fahd University, P.O. Box 1664, Al Khobar 31952, Saudi Arabia; CREDIMI FRE 2003 CNRS, University of Burgundy, Dijon, France; rboudershem@pmu.edu.sa

[†] Presented at The 11th International Electronic Conference on Sensors and Applications (ECSA-11), 26–28 November 2024; Available online: <https://sciforum.net/event/ecsa-11>

Abstract: This research proposes a comprehensive framework for implementing explainable and transparent artificial intelligence (AI) sensors in healthcare, addressing the challenges posed by AI “black boxes” while adhering to European Union (EU) AI Act and Data Act requirements. Our approach combines interpretable machine learning (ML), human-AI interaction, and ethical guidelines to ensure AI sensor outputs are comprehensible, auditable, and aligned with clinical decision-making. The framework consists of three core components. First, interpretable AI model architecture using techniques like attention mechanisms and symbolic reasoning. Second, an interactive interface facilitating collaboration between healthcare professionals and AI systems. Third, a robust ethical and regulatory framework addressing bias, privacy, and accountability. By tackling transparency and explainability challenges, our research aims to improve patient outcomes, support informed decision-making, and increase public acceptance of AI in healthcare. The proposed framework contributes to the responsible development of AI technologies in full compliance with EU regulations, ensuring alignment with the vision for trustworthy and human-centric AI systems. This approach paves the way for the safe and ethical adoption of AI sensors in healthcare, ultimately enhancing patient care while maintaining high standards of transparency and accountability.

Keywords: AI; explainability; privacy; public trust; sensors; transparency

1. Introduction

The rapid advancements in AI and ML have paved the way for transformative applications especially in healthcare delivery. AI-powered sensors and monitoring systems hold immense potential to revolutionize patient care by enabling early disease detection, personalized treatment plans, and continuous health monitoring [1,2]. However, the widespread adoption of AI in healthcare remains hindered by concerns over the opacity and lack of transparency in many AI systems, which can lead to issues of trust, accountability [3], and ethical implications [4]. In healthcare, where decisions can have profound impacts on human lives, it is crucial that AI systems are explainable and transparent [5], allowing healthcare professionals and patients to understand the reasoning behind their outputs and recommendations [6]. The opaque nature of many current AI models, often referred to as “black boxes” [7], poses significant challenges in terms of interpretability, fairness, and reliability, which are critical factors in healthcare applications [8]. The need for explainable and transparent AI (XAI) in healthcare has been widely acknowledged by researchers, practitioners, and policymakers. XAI aims to develop AI systems that are not only accurate and efficient [9] but also capable of providing human-understandable explanations for their decisions [10]. By making AI systems more interpretable and transparent, XAI can foster trust [11], enable effective human-AI collaboration, and facilitate the responsible deployment of AI in healthcare [12]. This research aims to address the challenges of developing explainable and transparent AI sensors for healthcare

Citation: Boudershem, R. A. Comprehensive Framework for Transparent and Explainable AI Sensors in Healthcare. *Eng. Proc.* **2024**, *6*, x. <https://doi.org/10.3390/xxxxx>

Academic Editor(s):

Published: date



Copyright: © 2024 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

applications. Specifically, we propose a comprehensive framework that integrates interpretable machine learning models, human-AI interaction mechanisms, and ethical guidelines to ensure that AI sensor outputs are comprehensible, auditable, and aligned with clinical decision-making processes. The proposed framework has three core components. Firstly, an interpretable AI model architecture that leverages techniques such as attention mechanisms [13], symbolic reasoning [14], and rule-based systems [15] to provide human-understandable explanations. Secondly, an interactive interface that facilitates effective communication and collaboration between healthcare professionals and AI systems [16], enabling seamless integration of AI insights into clinical workflows. Thirdly, a robust ethical and regulatory framework that addresses issues of bias [17], privacy [18], and accountability [19] in the deployment of AI sensors in healthcare. By developing explainable and transparent AI sensors tailored for healthcare applications, this research aims to contribute to the responsible development of AI technologies and pave the way for improved patient outcomes, informed decision-making, and increased public acceptance of AI in the healthcare domain [20]. Addressing the challenges of transparency and explainability is crucial for facilitating the safe and ethical adoption of AI sensors in healthcare.

2. Methodology

To develop a comprehensive framework for explainable and transparent AI sensors in healthcare, we employ a multi-pronged approach involving a systematic literature review and empirical analysis.

2.1. Comprehensive Literature Review

We conducted a comprehensive review of existing literature to identify the key requirements, challenges, and state-of-the-art techniques associated with developing transparent and explainable AI systems for healthcare applications. The literature review covered the following aspects.

2.1.1. Key Requirements and Challenges

We identified the critical factors for deploying AI systems in healthcare, such as interpretability [21], transparency, fairness, privacy, and accountability [22]. In addition, we examined the challenges and pitfalls of applying opaque “black-box” AI models in high-stakes healthcare situations [23].

2.1.2. Existing Approaches and Techniques

We explored various interpretable machine learning models and techniques, including attention mechanisms, symbolic reasoning, and rule-based systems. Then, we investigated human-AI interaction approaches for effective communication and collaboration between healthcare professionals and AI systems [24]. Finally, we analyzed ethical frameworks, guidelines, and regulatory considerations for responsible AI deployment in healthcare [25]. The literature review provided a solid foundation for understanding the current landscape, identifying ethical challenges, legal voids and informing the development of our proposed framework.

2.2. Empirical Analysis

To validate and refine our proposed framework, an empirical analysis involving data collection, preprocessing, and experimental evaluation is necessary and should consist of the following steps.

2.2.1. Data Collection and Preprocessing

First, we need to gather relevant healthcare datasets (e.g., electronic health records, sensor data, and medical images) from publicly available sources or collaborating healthcare institutions. PubMed, Web of Science and Scopus databases could also serve as

a starting point to collect relevant data. Second, we should preprocess the data to handle missing values, noise, and other data quality issues, while ensuring compliance with privacy and ethical guidelines.

2.2.2. Experimental Setup and Evaluation Metrics

The first step here is to implement and evaluate the components of our proposed framework, including interpretable AI models, interactive interfaces, ethical and regulatory considerations. The second step is to define appropriate evaluation metrics to assess the performance, interpretability, and transparency of our approach, such as predictive accuracy, model complexity, human-interpretability scores, and fairness measures so we can ensure data accuracy and relevance. The third step is to conduct controlled experiments and simulations to compare our framework with existing baseline methods and approaches. This empirical analysis will provide quantitative and qualitative insights into the effectiveness of our proposed framework, enabling further refinements and validating its real-world applicability in healthcare settings.

3. Proposed Framework

Building upon the insights gained from the literature review, we propose a comprehensive framework for developing explainable and transparent AI sensors in healthcare settings. The proposed framework consists of three core components.

3.1. Interpretable AI Model Architecture

To ensure that AI sensor outputs are comprehensible and explainable to healthcare professionals and patients, we leverage various interpretable machine learning techniques and model architectures (see Table 1).

Table 1. Interpretable AI model architecture.

Key Elements to Be Incorporated:

1. Attention mechanisms
 2. Symbolic reasoning
 3. Rule-based systems
 4. Human-understandable explanations
-

We employ attention mechanisms, which have proven effective in enhancing interpretability by highlighting the most relevant features or input regions contributing to model predictions [26]. Attention mechanisms enable the model to attend to the most salient aspects of the input data, facilitating human-understandable explanations. Incorporating symbolic reasoning techniques, such as inductive logic programming [27] and neuro-symbolic approaches [28], will allow our model to leverage logical rules and symbolic representations. This hybrid approach combines the reasoning capabilities of symbolic systems with the powerful pattern recognition abilities of neural networks, enabling more interpretable and explainable decision-making processes. Rule-based systems, which represent knowledge in the form of human-readable rules, can provide intuitive and transparent explanations for model outputs [29]. By integrating rule-based components into our model architecture, we aim to enhance the interpretability and auditability of AI sensor decisions, particularly in critical healthcare situations such as precision medicine. Our interpretable AI model architecture is designed to generate human-understandable explanations for its outputs, leveraging techniques such as local interpretable model-agnostic explanations (LIME) [30], SHapley Additive exPlanations (SHAP) [31], and counterfactual explanations [32]. These explanations can help healthcare professionals understand the reasoning behind AI sensor recommendations and facilitate effective human-AI collaboration.

3.2. Interactive Human-AI Interface

Effective communication and collaboration between healthcare professionals and AI systems are crucial for the successful integration of AI sensors into clinical workflows. To address this need, our framework incorporates an interactive human-AI interface that facilitates seamless human-AI interaction and decision-making (see Table 2).

Table 2. Interactive human-AI interface.

Key Aspects of the Interface:
1. Explanation visualization
2. Interactive querying
3. Collaborative workflow integration
4. User feedback and model refinement

The interface provides clear and intuitive visualizations of the explanations generated by the interpretable AI model, enabling healthcare professionals to understand the reasoning behind AI sensor outputs [33]. Interactive querying means that users can actively query the AI system, asking for clarifications, additional explanations, or alternative recommendations, fostering a collaborative decision-making process [34]. In addition, the interface is designed to seamlessly integrate AI sensor insights and recommendations into existing clinical workflows, minimizing disruptions and facilitating effective human-AI collaboration [35]. The interface incorporates mechanisms for healthcare professionals to provide feedback and annotate data, enabling continuous model refinement and improvement based on real-life clinical insights [36].

3.3. Ethical and Regulatory Framework

Deploying AI sensors in healthcare raises critical ethical and regulatory concerns, such as fairness, privacy, and accountability (see Table 3).

Table 3. Ethical and regulatory challenges.

Key Issues:
1. Bias mitigation, discrimination and fairness
2. Privacy and data protection
3. Accountability and auditing
4. Ethical guidelines and oversight
5. Transparency
6. Explainability
7. Performance
8. Data quality and accuracy
9. Cost-effectiveness and affordability
10. Errors and misdiagnosis
11. Access to health and technology for all

We employ techniques for detecting and mitigating biases in AI models, such as adversarial debiasing [37], causal reasoning [38], and fair representation learning [39]. These approaches aim to ensure fair and equitable AI sensor outputs, reducing the risk of discrimination or unfair treatment. Our framework also incorporates strong privacy-preserving measures, such as differential privacy [40], homomorphic encryption [41], and federated learning [42], to protect sensitive patient data and ensure compliance with relevant data protection regulations (e.g., HIPAA, GDPR, AI Act, Data Act). We implement mechanisms for auditing and documenting AI sensor decisions, model performance, and potential issues or failures [43]. This promotes accountability and enables thorough

investigation and remediation in case of adverse events or unforeseen consequences. Our framework adheres to established ethical guidelines and principles for AI in healthcare [44]. We also recommend the establishment of multidisciplinary oversight committees, including healthcare professionals, ethicists, patient advocates, and AI experts, to ensure responsible and ethical deployment of AI sensors.

By integrating these three core components—interpretable AI models, interactive human-AI interfaces, and ethical and regulatory frameworks—our proposed framework aims to facilitate the development and deployment of explainable and transparent AI sensors in healthcare settings, fostering trust, accountability, and responsible AI adoption.

3.4. Theoretical Framework and Research Hypotheses

Based on our comprehensive literature review and the proposed framework, we identify several key hypotheses regarding the expected performance and impact of our approach. These hypotheses will guide future empirical validation efforts and help establish the effective implementation of our framework in real-life situations for developing explainable AI systems in healthcare, such as traditional rule-based systems, post-hoc explanation techniques, and black-box models [45].

3.4.1. Hypotheses for Interpretable AI Model Architecture

According to our first assumption, the integration of attention mechanisms with symbolic reasoning will provide more interpretable explanations compared to traditional black-box models while maintaining comparable performance levels [46]. This hypothesis builds upon previous research demonstrating the effectiveness of attention mechanisms in neural networks and the interpretability advantages of symbolic systems.

Our second hypothesis suggests that the hybrid approach combining rule-based systems with machine learning will enable healthcare professionals to understand the reasoning behind AI recommendations more effectively than either approach alone. This hypothesis relies on the complementary nature of explicit rules and learned patterns in medical decision-making [47].

The third hypothesis proposes that the architecture will demonstrate adaptability across different healthcare situations while maintaining consistent levels of interpretability. This adaptability is crucial for the practical implementation of our framework across various medical specialties [48].

3.4.2. Human-AI Interface Hypotheses

Firstly, the interactive interface design will facilitate and improve collaboration between healthcare professionals and AI systems compared to traditional decision support systems. This improved collaboration will reduce decision-making time and enhanced quality of clinical decisions [49].

Secondly, real-time explanation capabilities will lead to increased trust and acceptance among healthcare professionals. This hypothesis addresses the critical role of transparency in building healthcare providers' confidence in AI-assisted decision-making [50].

Thirdly, the implementation of feedback mechanisms will enable continuous improvement of the system's performance and relevance in clinical settings. This continuous learning approach is essential for maintaining the effectiveness of the proposed framework over time and adapting to evolving clinical practices [51].

3.4.3. Ethical Framework Hypotheses

Firstly, the implementation of bias mitigation techniques will effectively reduce demographic disparities in AI outputs and recommendations across different patient populations, and consequently tend to reduce discrimination between patients, errors and

misdiagnosis. This hypothesis directly addresses concerns about fairness and equity in healthcare AI systems [52].

Secondly, implementing privacy-related safeguards and techniques will maintain system performance while ensuring compliance with data protection regulations such as the GDPR [53], the EU Data Act [54] or HIPAA [55]. This balance between functionality and privacy is crucial for practical implementation in healthcare delivery [56]. Moreover, it is important to note the convergence between the GDPR and the AI Act [57] regarding the protection of confidentiality and patient privacy.

Thirdly, the proposed accountability mechanisms will enable effective auditing and oversight of AI decision-making processes. This hypothesis addresses the growing need for responsible AI deployment in healthcare [58].

3.4.4. Proposed Validation Methodology

The formal validation of our hypotheses necessitates a holistic methodology combining both quantitative and qualitative approaches [59]. The quantitative validation will focus on measuring system performance, user interaction efficiency, and ethical compliance. This will include comparative analyses with existing systems such as the use of blockchain in healthcare delivery [60], assessment of prediction accuracy across healthcare situations, and evaluation of computational efficiency and scalability [61].

The qualitative validation will incorporate healthcare professional feedback through semi-structured interviews and observational studies. Expert panels comprising clinicians, ethicists, and technical specialists will review the framework's implementation and impact. These panels will assess clinical relevance, technical architecture, and regulatory compliance.

3.4.5. Expected Challenges and Implementation Considerations

This theoretical framework faces several anticipated challenges. Technical challenges include the complexity of integration with existing healthcare systems and the demands of real-time processing in clinical settings [62]. The framework must address variations in data quality [63] and standardization [64] across different healthcare institutions.

In addition, we can mention clinical challenges such as the adaptation of existing workflows and the training requirements for healthcare professionals [65]. The proposed framework must maintain consistency in its support for clinical decision-making and accommodate different practice patterns. Emergencies require particular attention to ensure the system remains helpful without impeding rapid response capabilities.

Organizational challenges include resource allocation for implementation and change management requirements [66]. The ratio cost-effectiveness of implementation and the need for policy and procedure updates to accommodate AI models in healthcare is an important factor for their universal adoption [67].

3.5. Future Validation Plan

Three distinct phases will determine the empirical validation of this framework. The initial phase will involve a pilot study within a single healthcare department, focusing on technical feasibility and gathering initial user feedback. This phase will establish baseline performance indicators and assess basic user acceptance.

The second phase will expand validation to multiple departments within one healthcare institution. This extended validation will provide comprehensive insights into performance across different clinical contexts and evaluate workflow integration effectiveness as well.

The final phase encompasses a multi-center validation across several healthcare institutions. This phase will assess the generalizability and scalability of our framework; here, we will examine performance and adaptation requirements across different healthcare systems and patient populations.

3.5.1. Expected Impact and Implications

The implementation of this framework will significantly influence clinical practice through enhanced decision-making support and improved patient care through early detection capabilities. Healthcare professionals will benefit from efficient clinical workflows and better documentation processes [68], leading to enhanced accountability in medical decision-making [69].

This framework will establish new standards for interpretability in medical AI systems. The proposed human-AI collaboration [70] model will advance our understanding of effective interaction between healthcare professionals and AI systems. We believe the privacy protection measures implemented in the proposed framework could not only set new benchmarks for securing sensitive medical data but also maintain system functionality and transparency [71].

3.5.2. Research Contributions

This theoretical framework proposes a novel integration of interpretable AI techniques specifically tailored to healthcare applications. As explained, the comprehensive human-AI interaction model deals with the unique requirements of clinical decision-making processes. In addition, the structured approach to ethical AI implementation provides a template for responsible deployment of AI systems in healthcare. Our hypotheses establish a foundation for future empirical research for a successful validation and implementation of this framework. Extensive collaboration among healthcare professionals, technical experts, and researchers is a requirement, as it will allow for empirical validation of the proposed hypotheses and refinement of the framework.

4. Discussion and Future Directions

4.1. *Enhancing Transparency, Explainability, and Trust*

4.1.1. Interpretability of AI Sensor Outputs

Our interpretable AI model architecture could have the ability to provide human-understandable explanations for AI sensor outputs, enhancing transparency and facilitating trust between healthcare professionals and AI systems [72]. The attention mechanisms, symbolic reasoning, and rule-based components could significantly contribute to the interpretability of the model, enabling healthcare professionals to understand the reasoning behind recommendations and decisions.

4.1.2. Healthcare Professional-AI Collaboration

The interactive human-AI interface will facilitate effective communication and collaboration between healthcare professionals and AI systems, enabling a seamless integration of AI sensor insights into clinical workflows [73]. User feedback and model refinement mechanisms also allow for continuous improvement and adaptation of the AI models based on real-world clinical insights and experiences.

4.1.3. Addressing Ethical and Regulatory Concerns

Our ethical and regulatory framework will effectively mitigate biases in AI sensor outputs, reducing the risk of unfair treatment or discrimination against certain patient groups [74]. Strong privacy-preserving measures and data protection techniques will ensure compliance with relevant regulations and protect sensitive patient data from potential privacy attacks or breaches. Accountability and auditing mechanisms will enable thorough investigation and remediation in case of adverse events or unforeseen consequences, promoting responsible AI deployment in healthcare settings.

4.2. Limitations and Future Research Directions

While our proposed framework highlights promising results in developing explainable and transparent AI sensors for healthcare, we acknowledge several limitations and outline potential future research directions.

4.2.1. Scalability and Computational Complexity

Some components of our framework, such as attention mechanisms and symbolic reasoning, may introduce additional computational complexity, which could pose challenges when scaling to large healthcare datasets or real-time applications [75]. Future research should explore efficient implementations and optimizations to address scalability and computational resource requirements.

4.2.2. Generalizability Across Healthcare Domains

Further evaluation is necessary to assess the generalizability of our framework across diverse healthcare domains, such as radiology, genomics, and mental health [76].

4.2.3. Continuous Model Refinement and Adaptation

As healthcare practices and regulations evolve, there is a need for mechanisms to continuously refine and adapt our AI models and frameworks to stay relevant and compliant [77]. Integrating techniques for continual learning, transfer learning, and domain adaptation could enhance the long-term applicability and adaptability of our framework.

4.2.4. Integrating Multi-Modal Data Sources

Many healthcare applications involve multi-modal data sources, such as electronic health records, medical images, sensor data, and genomic information [78]. Future research should explore methods to effectively integrate and interpret multi-modal data sources within our framework, enabling more comprehensive and holistic AI-driven healthcare solutions.

4.2.5. Fostering Trust and Acceptance

Despite the efforts to enhance transparency and explainability, building trust and acceptance among healthcare professionals, patients, and the public remains a significant challenge [79]. Interdisciplinary collaborations, public education initiatives, and stakeholder engagement are crucial to address this challenge and facilitate responsible AI adoption in healthcare. The safe and ethical deployment of explainable and transparent AI sensors in healthcare settings is subject to numerous challenges and has to be constantly monitored and improved; by adopting such approach, we will contribute to improve patient outcomes and informed decision-making.

5. Conclusions

The responsible development and deployment of AI technologies, particularly in high-stakes domains like healthcare, is of paramount importance. Our research contributes to this goal by providing a comprehensive framework that prioritizes transparency, explainability, and ethical considerations throughout the AI development lifecycle. By making AI systems more interpretable and facilitating human-AI collaboration, our approach empowers healthcare professionals to understand and trust the reasoning behind AI-driven recommendations and decisions. This trust is crucial for the successful adoption and integration of AI technologies in healthcare settings, ultimately contributing to improved patient outcomes and informed decision-making processes.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Publicly available data.

Conflicts of Interest: The author declares no conflict of interest.

References

1. Topol, E.J. High-performance medicine: The convergence of human and artificial intelligence. *Nat. Med.* **2019**, *25*, 44–56. <https://doi.org/10.1038/s41591-018-0300-7>.
2. Nzomo, M.; Moodley, D. A Semantic Architecture for Continuous Health Monitoring, Risk Prediction, and Proactive Decision Making. In *AI for Health Equity and Fairness. Studies in Computational Intelligence*; Shaban-Nejad, A., Michalowski, M., Bianco, S., Eds.; Springer: Cham, Switzerland, 2024; Volume 1164. https://doi.org/10.1007/978-3-031-63592-2_20.
3. Smith, H. Clinical AI: Opacity, accountability, responsibility and liability. *AI Soc.* **2021**, *36*, 535–545. <https://doi.org/10.1007/s00146-020-01019-6>.
4. Elendu, C.; Amaechi, D.C.M.; Elendu, T.C.B.; Jingwa, K.A.M.; Okoye, O.K.M.; Okah, M.M.J.; Ladele, J.A.M.; Farah, A.H.; Alimi, H.A.M. Ethical implications of AI and robotics in healthcare: A review. *Medicine* **2023**, *102*, e36671. <https://doi.org/10.1097/md.00000000000036671>.
5. Hulsén, T. Explainable Artificial Intelligence (XAI): Concepts and Challenges in Healthcare. *AI* **2023**, *4*, 652–666. <https://doi.org/10.3390/ai4030034>.
6. Johnson, K.B.; Wei, W.; Weeraratne, D.; Frisse, M.E.; Misulis, K.; Rhee, K.; Zhao, J.; Snowdon, J.L. Precision Medicine, AI, and the Future of Personalized Health Care. *Clin. Transl. Sci.* **2021**, *14*, 86–93. <https://doi.org/10.1111/cts.12884>.
7. Wadden, J.J. Defining the undefinable: The black box problem in healthcare artificial intelligence. *J. Med. Ethic.* **2022**, *48*, 764–768. <https://doi.org/10.1136/medethics-2021-107529>.
8. Valente, F.; Paredes, S.; Henriques, J.; Rocha, T.; de Carvalho, P.; Morais, J. Interpretability, personalization and reliability of a machine learning based clinical decision support system. *Data Min. Knowl. Discov.* **2022**, *36*, 1140–1173. <https://doi.org/10.1007/s10618-022-00821-8>.
9. Manresa-Yee, C.; Roig-Maimó, M.F.; Ramis, S.; Mas-Sansó, R. Advances in XAI: Explanation Interfaces in Healthcare. In *Handbook of Artificial Intelligence in Healthcare*; Lim, C.P., Chen, Y.W., Vaidya, A., Mahorkar, C., Jain, L.C., Eds.; Intelligent Systems Reference Library; Springer: Cham, Switzerland, 2022; Volume 212. https://doi.org/10.1007/978-3-030-83620-7_15.
10. Chaddad, A.; Peng, J.; Xu, J.; Bouridane, A. Survey of Explainable AI Techniques in Healthcare. *Sensors* **2023**, *23*, 634. <https://doi.org/10.3390/s23020634>.
11. Gerlings, J.; Jensen, M.S.; Shollo, A. Explainable AI, But Explainable to Whom? An Exploratory Case Study of xAI in Healthcare. In *Handbook of Artificial Intelligence in Healthcare*; Lim, C.P., Chen, Y.W., Vaidya, A., Mahorkar, C., Jain, L.C., Eds.; Intelligent Systems Reference Library; Springer: Cham, Switzerland, 2022; Volume 212. https://doi.org/10.1007/978-3-030-83620-7_7.
12. Akhtar, M.A.K.; Kumar, M.; Nayyar, A. Socially Responsible Applications of Explainable AI. In *Towards Ethical and Socially Responsible Explainable AI*; Studies in Systems, Decision and Control; Springer: Cham, Switzerland, 2024; Volume 551. https://doi.org/10.1007/978-3-031-66489-2_9.
13. Rajabi, E.; Kafaie, S. Knowledge Graphs and Explainable AI in Healthcare. *Information* **2022**, *13*, 459. <https://doi.org/10.3390/info13100459>.
14. Van Woensel, W.; Scioscia, F.; Loseto, G.; Seneviratne, O.; Patton, E.; Abidi, S. Explanations of Symbolic Reasoning to Effect Patient Persuasion and Education. In *Explainable Artificial Intelligence and Process Mining Applications for Healthcare*. Juarez, J.M., et al., Ed.; XAI-Healthcare PM4H 2023 2023; Communications in Computer and Information Science; Springer: Cham, Switzerland, 2024; Volume 2020. https://doi.org/10.1007/978-3-031-54303-6_7.
15. Kim, S.Y.; Kim, D.H.; Kim, M.J.; Ko, H.J.; Jeong, O.R. XAI-Based Clinical Decision Support Systems: A Systematic Review. *Appl. Sci.* **2024**, *14*, 6638. <https://doi.org/10.3390/app14156638>.
16. Petersson, L.; Larsson, I.; Nygren, J.M.; Nilsen, P.; Neher, M.; Reed, J.E.; Tyskbo, D.; Svedberg, P. Challenges to implementing artificial intelligence in healthcare: A qualitative interview study with healthcare leaders in Sweden. *BMC Health Serv. Res.* **2022**, *22*, 850. <https://doi.org/10.1186/s12913-022-08215-8>.
17. Kerasidou, A. Ethics of artificial intelligence in global health: Explainability, algorithmic bias and trust. *J. Oral Biol. Craniofacial Res.* **2021**, *11*, 612–614. <https://doi.org/10.1016/j.jobcr.2021.09.004>.
18. Khalid, N.; Qayyum, A.; Bilal, M.; Al-Fuqaha, A.; Qadir, J. Privacy-preserving artificial intelligence in healthcare: Techniques and applications. *Comput. Biol. Med.* **2023**, *158*, 106848. <https://doi.org/10.1016/j.compbiomed.2023.106848>.
19. Shaban-Nejad, A.; Michalowski, M.; Brownstein, J.S.; Buckeridge, D.L. Guest Editorial Explainable AI: Towards Fairness, Accountability, Transparency and Trust in Healthcare. *IEEE J. Biomed. Health Inform.* **2021**, *25*, 2374–2375. <https://doi.org/10.1109/jbhi.2021.3088832>.
20. Liu, K.; Tao, D. The roles of trust, personalization, loss of privacy, and anthropomorphism in public acceptance of smart healthcare services. *Comput. Hum. Behav.* **2021**, *127*, 107026. <https://doi.org/10.1016/j.chb.2021.107026>.
21. MacDonald, S.; Steven, K.; Trzaskowski, M. Interpretable AI in Healthcare: Enhancing Fairness, Safety, and Trust. In *Artificial Intelligence in Medicine*; Raz, M., Nguyen, T.C., Loh, E., Eds.; Springer: Singapore, 2022. https://doi.org/10.1007/978-981-19-1223-8_11.

22. Ueda, D.; Kakinuma, T.; Fujita, S.; Kamagata, K.; Fushimi, Y.; Ito, R.; Matsui, Y.; Nozaki, T.; Nakaura, T.; Fujima, N.; et al. Fairness of artificial intelligence in healthcare: Review and recommendations. *Jpn. J. Radiol.* **2024**, *42*, 3–15. <https://doi.org/10.1007/s11604-023-01474-3>.
23. Rudin, C. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nat. Mach. Intell.* **2019**, *1*, 206–215. <https://doi.org/10.1038/s42256-019-0048-x>.
24. Abedin, B.; Meske, C.; Junglas, I.; Rabhi, F.; Motahari-Nezhad, H.R. Designing and Managing Human-AI Interactions. *Inf. Syst. Front.* **2022**, *24*, 691–697. <https://doi.org/10.1007/s10796-022-10313-1>.
25. Siala, H.; Wang, Y. SHIFTing artificial intelligence to be responsible in healthcare: A systematic review. *Soc. Sci. Med.* **2022**, *296*, 114782. <https://doi.org/10.1016/j.socscimed.2022.114782>.
26. Park, S.; Koh, Y.; Jeon, H.; Kim, H.; Yeo, Y.; Kang, J. Enhancing the interpretability of transcription factor binding site prediction using attention mechanism. *Sci. Rep.* **2020**, *10*, 13413. <https://doi.org/10.1038/s41598-020-70218-4>.
27. Calegari, R.; Ciatto, G.; Denti, E.; Omicini, A. Logic-Based Technologies for Intelligent Systems: State of the Art and Perspectives. *Information* **2020**, *11*, 167. <https://doi.org/10.3390/info11030167>.
28. Lu, Z.; Afridi, I.; Kang, H.J.; Ruchkin, I.; Zheng, X. Surveying neuro-symbolic approaches for reliable artificial intelligence of things. *J. Reliab. Intell. Environ.* **2024**, *10*, 257–279. <https://doi.org/10.1007/s40860-024-00231-1>.
29. Hassija, V.; Chamola, V.; Mahapatra, A.; Singal, A.; Goel, D.; Huang, K.; Scardapane, S.; Spinelli, I.; Mahmud, M.; Hussain, A. Interpreting Black-Box Models: A Review on Explainable Artificial Intelligence. *Cogn. Comput.* **2024**, *16*, 45–74. <https://doi.org/10.1007/s12559-023-10179-8>.
30. Band, S.S.; Yarahmadi, A.; Hsu, C.-C.; Biyari, M.; Sookhak, M.; Ameri, R.; Dehzangi, I.; Chronopoulos, A.T.; Liang, H.-W. Application of explainable artificial intelligence in medical health: A systematic review of interpretability methods. *Inform. Med. Unlocked* **2023**, *40*, 101286. <https://doi.org/10.1016/j.imu.2023.101286>.
31. Guleria, P.; Srinivasu, P.N.; Hassaballah, M. Diabetes prediction using Shapley additive explanations and DSaaS over machine learning classifiers: A novel healthcare paradigm. *Multimed. Tools Appl.* **2023**, *83*, 40677–40712. <https://doi.org/10.1007/s11042-023-17212-w>.
32. Durán, J.M. Dissecting scientific explanation in AI (sXAI): A case for medicine and healthcare. *Artif. Intell.* **2021**, *297*, 103498. <https://doi.org/10.1016/j.artint.2021.103498>.
33. Ooge, J.; Stiglic, G.; Verbert, K. Explaining artificial intelligence with visual analytics in healthcare. *WIREs Data Min. Knowl. Discov.* **2022**, *12*, e1427. <https://doi.org/10.1002/widm.1427>.
34. Vellido, A. The importance of interpretability and visualization in machine learning for applications in medicine and health care. *Neural Comput. Appl.* **2019**, *32*, 18069–18083. <https://doi.org/10.1007/s00521-019-04051-w>.
35. Nasarian, E.; Alizadehsani, R.; Acharya, U.; Tsui, K.-L. Designing interpretable ML system to enhance trust in healthcare: A systematic review to proposed responsible clinician-AI-collaboration framework. *Inf. Fusion* **2024**, *108*, 102412. <https://doi.org/10.1016/j.inffus.2024.102412>.
36. Payrovnaziri, S.N.; Chen, Z.; Rengifo-Moreno, P.; Miller, T.; Bian, J.; Chen, J.H.; Liu, X.; He, Z. Explainable artificial intelligence models using real-world electronic health record data: A systematic scoping review. *J. Am. Med. Inform. Assoc.* **2020**, *27*, 1173–1185. <https://doi.org/10.1093/jamia/ocaa053>.
37. Ferrara, E. Fairness and Bias in Artificial Intelligence: A Brief Survey of Sources, Impacts, and Mitigation Strategies. *Sci* **2024**, *6*, 3. <https://doi.org/10.3390/sci6010003>.
38. Ntoutsis, E.; Fafalios, P.; Gadiraju, U.; Iosifidis, V.; Nejdil, W.; Vidal, M.; Ruggieri, S.; Turini, F.; Papadopoulos, S.; Krasanakis, E.; et al. Bias in data-driven artificial intelligence systems—An introductory survey. *WIREs Data Min. Knowl. Discov.* **2020**, *10*, e1356. <https://doi.org/10.1002/widm.1356>.
39. Hort, M.; Chen, Z.; Zhang, J.M.; Harman, M.; Sarro, F. Bias Mitigation for Machine Learning Classifiers: A Comprehensive Survey. *Acm J. Responsible Comput.* **2024**, *1*, 1–52. <https://doi.org/10.1145/3631326>.
40. Giuffrè, M.; Shung, D.L. Harnessing the power of synthetic data in healthcare: Innovation, application, and privacy. *npj Digit. Med.* **2023**, *6*, 186. <https://doi.org/10.1038/s41746-023-00927-3>.
41. Ali, A.; Pasha, M.F.; Ali, J.; Fang, O.H.; Masud, M.; Jurcut, A.D.; Alzain, M.A. Deep Learning Based Homomorphic Secure Search-Able Encryption for Keyword Search in Blockchain Healthcare System: A Novel Approach to Cryptography. *Sensors* **2022**, *22*, 528. <https://doi.org/10.3390/s22020528>.
42. Rahman, A.; Hossain, S.; Muhammad, G.; Kundu, D.; Debnath, T.; Rahman, M.; Khan, S.I.; Tiwari, P.; Band, S.S. Federated learning-based AI approaches in smart healthcare: Concepts, taxonomies, challenges and open issues. *Clust. Comput.* **2023**, *26*, 2271–2311. <https://doi.org/10.1007/s10586-022-03658-4>.
43. Falco, G.; Shneiderman, B.; Badger, J.; Carrier, R.; Dahbura, A.; Danks, D.; Eling, M.; Goodloe, A.; Gupta, J.; Hart, C.; et al. Governing AI safety through independent audits. *Nat. Mach. Intell.* **2021**, *3*, 566–571. <https://doi.org/10.1038/s42256-021-00370-7>.
44. Oniani, D.; Hilsman, J.; Peng, Y.; Poropatich, R.K.; Pamplin, J.C.; Legault, G.L.; Wang, Y. Adopting and expanding ethical principles for generative artificial intelligence from military to healthcare. *npj Digit. Med.* **2023**, *6*, 225. <https://doi.org/10.1038/s41746-023-00965-x>.
45. Srinivasu, P.N.; Sandhya, N.; Jhaveri, R.H.; Raut, R. From Blackbox to Explainable AI in Healthcare: Existing Tools and Case Studies. *Mob. Inf. Syst.* **2022**, *2022*, 167821. <https://doi.org/10.1155/2022/8167821>.

46. Zhang, J.; Chen, B.; Zhang, L.; Ke, X.; Ding, H. Neural, symbolic and neural-symbolic reasoning on knowledge graphs. *AI Open* **2021**, *2*, 14–35. <https://doi.org/10.1016/j.aiopen.2021.03.001>.
47. Kierner, S.; Kucharski, J.; Kierner, Z. Taxonomy of hybrid architectures involving rule-based reasoning and machine learning in clinical decision systems: A scoping review. *J. Biomed. Inform.* **2023**, *144*, 104428. <https://doi.org/10.1016/j.jbi.2023.104428>.
48. Vellido, A. The importance of interpretability and visualization in machine learning for applications in medicine and health care. *Neural Comput. Appl.* **2019**, *32*, 18069–18083. <https://doi.org/10.1007/s00521-019-04051-w>.
49. Nazar, M.; Alam, M.M.; Yafi, E.; Su'Ud, M.M. A Systematic Review of Human–Computer Interaction and Explainable Artificial Intelligence in Healthcare With Artificial Intelligence Techniques. *IEEE Access* **2021**, *9*, 153316–153348. <https://doi.org/10.1109/access.2021.3127881>.
50. Asan, O.; Bayrak, A.E.; Choudhury, A. Artificial Intelligence and Human Trust in Healthcare: Focus on Clinicians. *J. Med. Internet Res.* **2020**, *22*, e15154. <https://doi.org/10.2196/15154>.
51. Kiyasseh, D.; Laca, J.; Haque, T.F.; Miles, B.J.; Wagner, C.; Donoho, D.A.; Anandkumar, A.; Hung, A.J. A multi-institutional study using artificial intelligence to provide reliable and fair feedback to surgeons. *Commun. Med.* **2023**, *3*, 42. <https://doi.org/10.1038/s43856-023-00263-3>.
52. Nazer, L.H.; Zatarah, R.; Waldrip, S.; Ke, J.X.C.; Moukheiber, M.; Khanna, A.K.; Hicklen, R.S.; Moukheiber, L.; Moukheiber, D.; Ma, H.; et al. Bias in artificial intelligence algorithms and recommendations for mitigation. *PLOS Digit. Health* **2023**, *2*, e0000278. <https://doi.org/10.1371/journal.pdig.0000278>.
53. Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the Protection of Natural Persons with Regard to the Processing of Personal Data and on the Free Movement of Such Data, and Repealing Directive 95/46/EC (General Data Protection Regulation) (Text with EEA relevance) (OJ L 119 04.05.2016, p. 1, ELI). Available online: <http://data.europa.eu/eli/reg/2016/679/oj> (accessed on).
54. Regulation (EU) 2023/2854 of the European Parliament and of the Council of 13 December 2023 on harmonised rules on fair access to and use of data and amending Regulation (EU) 2017/2394 and Directive (EU) 2020/1828 (Data Act) (Text with EEA relevance), PE/49/2023/REV/1, OJ L, 2023/2854, 22.12.2023, ELI. Available online: <http://data.europa.eu/eli/reg/2023/2854/oj> (accessed on).
55. Edemekong, P.; Annamaraju, P.; Haydel, M.J. *Health Insurance Portability and Accountability Act*; Updated 12 February 2024; StatPearls Publishing: Treasure Island, FL, USA; Available online: <https://www.ncbi.nlm.nih.gov/books/NBK500019/> (accessed on).
56. Williamson, S.M.; Prybutok, V. Balancing Privacy and Progress: A Review of Privacy Challenges, Systemic Oversight, and Patient Perceptions in AI-Driven Healthcare. *Appl. Sci.* **2024**, *14*, 675. <https://doi.org/10.3390/app14020675>.
57. Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 Laying down Harmonised Rules on Artificial Intelligence and AMENDING REGULATIONS (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act) (Text with EEA relevance), PE/24/2024/REV/1, OJ L, 2024/1689, 12.7.2024, ELI. Available online: <http://data.europa.eu/eli/reg/2024/1689/oj> (accessed on).
58. Inioluwa Deborah Raji, Andrew Smart, Rebecca, N. White, Margaret Mitchell, Timnit Gebru, Ben Hutchinson, Jamila Smith-Loud, Daniel Theron, and Parker Barnes. 2020. Closing the AI accountability gap: Defining an end-to-end framework for internal algorithmic auditing. In Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency (FAT* '20). Association for Computing Machinery, New York, NY, USA, 27–30 January 2020; pp. 33–44. <https://doi.org/10.1145/3351095.3372873>.
59. Secinaro, S.; Calandra, D.; Secinaro, A.; Muthurangu, V.; Biancone, P. The role of artificial intelligence in healthcare: A structured literature review. *BMC Med. Inform. Decis. Mak.* **2021**, *21*, 125. <https://doi.org/10.1186/s12911-021-01488-9>.
60. Sai, S.; Chamola, V.; Choo, K.-K.R.; Sikdar, B.; Rodrigues, J.J.P.C. Confluence of Blockchain and Artificial Intelligence Technologies for Secure and Scalable Healthcare Solutions: A Review. *IEEE Internet Things J.* **2022**, *10*, 5873–5897. <https://doi.org/10.1109/jiot.2022.3232793>.
61. Ali, A.; Ali, H.; Saeed, A.; Khan, A.A.; Tin, T.T.; Assam, M.; Ghadi, Y.Y.; Mohamed, H.G. Blockchain-Powered Healthcare Systems: Enhancing Scalability and Security with Hybrid Deep Learning. *Sensors* **2023**, *23*, 7740. <https://doi.org/10.3390/s23187740>.
62. Reddy, S.; Fox, J.; Purohit, M.P. Artificial intelligence-enabled healthcare delivery. *J. R. Soc. Med.* **2018**, *112*, 22–28. <https://doi.org/10.1177/0141076818815510>.
63. Pezoulas, V.C.; Kourou, K.D.; Kalatzis, F.; Exarchos, T.P.; Venetsanopoulou, A.; Zampeli, E.; Gandolfo, S.; Skopouli, F.; De Vita, S.; Tzioufas, A.G.; et al. Medical data quality assessment: On the development of an automated framework for medical data curation. *Comput. Biol. Med.* **2019**, *107*, 270–283. <https://doi.org/10.1016/j.compbimed.2019.03.001>.
64. Hoc Group on Application of AI Technologies. Artificial Intelligence in Healthcare: Directions of Standardization. In *Handbook of Artificial Intelligence in Healthcare*; Lim, C.P., Chen, Y.W., Vaidya, A., Mahorkar, C., Jain, L.C., Eds.; Intelligent Systems Reference Library; Springer: Cham, Switzerland, 2022; Volume 212. https://doi.org/10.1007/978-3-030-83620-7_10.
65. Esmailzadeh, P. Challenges and strategies for wide-scale artificial intelligence (AI) deployment in healthcare practices: A perspective for healthcare organizations. *Artif. Intell. Med.* **2024**, *151*, 102861. <https://doi.org/10.1016/j.artmed.2024.102861>.
66. Formosa, P.; Rogers, W.; Griep, Y.; Bankins, S.; Richards, D. Medical AI and human dignity: Contrasting perceptions of human and artificially intelligent (AI) decision making in diagnostic and medical resource allocation contexts. *Comput. Hum. Behav.* **2022**, *133*, 107296. <https://doi.org/10.1016/j.chb.2022.107296>.

67. Zhang, Y.; Hu, Y.; Jiang, N.; Yetisen, A.K. Wearable artificial intelligence biosensor networks. *Biosens. Bioelectron.* **2022**, *219*, 114825. <https://doi.org/10.1016/j.bios.2022.114825>.
68. Pateraki, M.; Fysarakis, K.; Sakkalis, V.; Spanoudakis, G.; Varlamis, I.; Maniadakis, M.; Lourakis, M.; Ioannidis, S.; Cummins, N.; Schuller, B.; et al. Chapter 2—Biosensors and Internet of Things in smart healthcare applications: Challenges and opportunities. In *Advances in ubiquitous sensing applications for healthcare, Wearable and Implantable Medical Devices*; Nilanjan, D., Amira, S.A., Simon, J.F., Chintan, B., Eds.; Academic Press: Cambridge, MA, USA, 2020, Volume 7, pp. 25–53, ISBN 9780128153697. <https://doi.org/10.1016/B978-0-12-815369-7.00002-1>.
69. Manickam, P.; Mariappan, S.A.; Murugesan, S.M.; Hansda, S.; Kaushik, A.; Shinde, R.; Thipperudraswamy, S.P. Artificial Intelligence (AI) and Internet of Medical Things (IoMT) Assisted Biomedical Systems for Intelligent Healthcare. *Biosensors* **2022**, *12*, 562. <https://doi.org/10.3390/bios12080562>.
70. Rawas, S. AI: The future of humanity. *Discov. Artif. Intell.* **2024**, *4*, 25. <https://doi.org/10.1007/s44163-024-00118-3>.
71. Kiseleva, A.; Kotzinos, D.; De Hert, P. Transparency of AI in Healthcare as a Multilayered System of Accountabilities: Between Legal Requirements and Technical Limitations. *Front. Artif. Intell.* **2022**, *5*, 879603. <https://doi.org/10.3389/frai.2022.879603>.
72. Ahmed, M.; Zubair, S. Explainable Artificial Intelligence in Sustainable Smart Healthcare. In *Explainable Artificial Intelligence for Cyber Security*; Ahmed, M., Islam, S.R., Anwar, A., Moustafa, N., Pathan, A.S.K., Eds.; Studies in Computational Intelligence; Springer: Cham, Switzerland, 2022; Volume 1025. https://doi.org/10.1007/978-3-030-96630-0_12.
73. Chen, E.; Prakash, S.; Reddi, V.J.; Kim, D.; Rajpurkar, P. A framework for integrating artificial intelligence for clinical care with continuous therapeutic monitoring. *Nat. Biomed. Eng.* **2023**, 1–10. <https://doi.org/10.1038/s41551-023-01115-0>.
74. Chen, R.J.; Wang, J.J.; Williamson, D.F.K.; Chen, T.Y.; Lipkova, J.; Lu, M.Y.; Sahai, S.; Mahmood, F. Algorithmic fairness in artificial intelligence for medicine and healthcare. *Nat. Biomed. Eng.* **2023**, *7*, 719–742. <https://doi.org/10.1038/s41551-023-01056-8>.
75. Aminizadeh, S.; Heidari, A.; Dehghan, M.; Toumaj, S.; Rezaei, M.; Navimipour, N.J.; Stroppa, F.; Unal, M. Opportunities and challenges of artificial intelligence and distributed systems to improve the quality of healthcare service. *Artif. Intell. Med.* **2024**, *149*, 102779. <https://doi.org/10.1016/j.artmed.2024.102779>.
76. Koutsouleris, N.; Hauser, T.U.; Skvortsova, V.; De Choudhury, M. From promise to practice: Towards the realisation of AI-informed mental health care. *Lancet Digit. Health* **2022**, *4*, e829–e840. [https://doi.org/10.1016/s2589-7500\(22\)00153-4](https://doi.org/10.1016/s2589-7500(22)00153-4).
77. Xia, B.; Lu, Q.; Zhu, L.; Lee, S.U.; Liu, Y.; Xing, Z. Towards a Responsible AI Metrics Catalogue: A Collection of Metrics for AI Accountability. In Proceedings of the IEEE/ACM 3rd International Conference on AI Engineering—Software Engineering for AI (CAIN '24). Association for Computing Machinery, New York, NY, USA, 14–15 April 2024; pp. 100–111. <https://doi.org/10.1145/3644815.3644959>.
78. Huang, S.-C.; Pareek, A.; Seyyedi, S.; Banerjee, I.; Lungren, M.P. Fusion of medical imaging and electronic health records using deep learning: A systematic review and implementation guidelines. *NPJ Digit. Med.* **2020**, *3*, 136. <https://doi.org/10.1038/s41746-020-00341-z>.
79. Kim, S.D. Application and Challenges of the Technology Acceptance Model in Elderly Healthcare: Insights from ChatGPT. *Technologies* **2024**, *12*, 68. <https://doi.org/10.3390/technologies12050068>.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.