Fostering Physician Trust: The Critical Role of Explainable AI in Clinical Decision Making

Rasit Dinc

Rasit Dinc Digital Health & AI Research

Published: July 12, 2025 | AI Diagnostics

DOI: 10.5281/zenodo.17996633

Abstract

The integration of Artificial Intelligence AI into healthcare has ushered in an era of unprecedented diagnostic and prognostic power. Modern AI models, parti...

The integration of Artificial Intelligence (AI) into healthcare has ushered in an era of unprecedented diagnostic and prognostic power. Modern AI models, particularly those leveraging deep learning, can analyze complex patient data—from medical images to genomic sequences—with a performance that often rivals or exceeds human experts. However, this high performance frequently comes at the cost of transparency. These sophisticated algorithms often operate as "black boxes," obscuring the reasoning behind their outputs. For a field as high-stakes as medicine, where every decision carries profound ethical and legal weight, this lack of clarity presents a major barrier to widespread adoption and, critically, to **physician trust**. Explainable AI (XAI) is not merely a technical feature; it is the essential bridge required to integrate AI successfully and ethically into the clinical workflow [1].

The Imperative for Explainability in Medicine

In contrast to many other industries, the application of AI in healthcare demands more than just predictive accuracy. A physician cannot simply accept an AI-generated recommendation without understanding the underlying rationale. This need for interpretability stems from several core clinical and ethical responsibilities.

Firstly, physicians bear the ultimate responsibility for patient care. To maintain accountability, they must be able to **validate the AI's output** and ensure it is based on clinically sound evidence, not spurious correlations or data biases. Secondly, the physician must be able to justify the decision to the patient, maintaining the principle of informed consent. Without an explanation, the AI’s recommendation is reduced to an opaque suggestion, undermining the physician-patient relationship and the legal requirement for due diligence [2].

XAI transforms AI-powered Clinical Decision Support Systems (CDSS) from mere suggestion engines into trustworthy, collaborative tools. By providing

clear, concise, and clinically relevant explanations, XAI ensures that the AI serves as a partner in decision-making, not a replacement for human judgment.

XAI as the Foundation of Physician Trust

Trust in a clinical context is not blind acceptance; it is a confidence built on understanding and verification. For physicians, this means having the necessary information to assess the reliability and validity of an AI’s output.

Research has consistently shown that XAI significantly increases clinicians' trust compared to standard, uninterpretable AI models [3]. This is particularly true when the explanations are tailored to the clinical context, such as highlighting the specific features in a medical image or the patient data points that most influenced the prediction. The "black box" nature of non-interpretable models, such as complex deep neural networks, remains a significant challenge to overcome, despite their superior performance in certain tasks [4]. XAI methods, including model-agnostic techniques like LIME (Local Interpretable Model-agnostic Explanations) and SHAP (SHapley Additive exPlanations), are vital for unveiling the inner workings of these complex models.

A key practical benefit of XAI is its ability to facilitate the detection of model errors and biases. For instance, an XAI system might reveal that a diagnostic model is relying on non-clinical data, such as the hospital's logo or a specific artifact in an image, rather than the actual pathology. Identifying these **spurious correlations** is crucial for preventing misdiagnosis and builds confidence in the model's true clinical relevance when it is operating correctly. This transparency allows for a necessary feedback loop, enabling developers and clinicians to refine models and ensure they are robust and fair.

Challenges and Future Directions

Despite its promise, the integration of XAI into clinical practice faces several challenges. Current XAI methods are often designed by computer scientists and may not be sufficiently geared toward the specific needs and workflows of clinicians [5]. The explanations provided can sometimes be too technical, abstract, or lack the necessary context to be actionable in a fast-paced clinical setting. Furthermore, there is an inherent trade-off between model performance and explainability; highly accurate, complex models are often the least interpretable.

The future of XAI in healthcare lies in a **human-centered design approach**. This requires developing XAI systems that provide explanations which are not only technically sound but also: 1. **Actionable:** The explanation should suggest a clear course of action or validation. 2. **Context-Aware:** The explanation must be relevant to the specific patient, clinical setting, and physician's expertise. 3. **Seamlessly Integrated:** The XAI output should be presented within the existing Electronic Health Record (EHR) or CDSS interface without disrupting the clinical workflow.

Conclusion

Explainable AI is not a mere technological enhancement; it is a fundamental requirement for the ethical, legal, and practical adoption of AI in clinical decision-making. By prioritizing transparency and interpretability, the digital health community can effectively dismantle the "black box" barrier. XAI ensures that AI functions as a trusted, accountable partner, empowering physicians to maintain their professional autonomy and ultimately leading to safer, more effective patient care.

**

References

[1] Abbas, Q. (2025). [Explainable AI in Clinical Decision Support Systems] (https://pmc.ncbi.nlm.nih.gov/articles/PMC12427955/). NCBI PMC*. [2] Amann, J. (2020). [Explainability for artificial intelligence in healthcare] (https://bmcmedinformdecismak.biomedcentral.com/articles/10.1186/s12911-020-01332-6).* BMC Medical Informatics and Decision Making*. [3] Rosenbacke, R. (2024). [How Explainable Artificial Intelligence Can Increase Clinicians' Trust](https://ai.jmir.org/2024/1/e53207).* JMIR AI*. [4] Noor, AA. (2025). [Unveiling Explainable AI in Healthcare: Current Trends] (https://wires.onlinelibrary.wiley.com/doi/full/10.1002/widm.70018).* WIRES Data Mining and Knowledge Discovery*. [5] Räz, T. (2025). [Explainable AI in medicine: challenges of integrating XAI into clinical practice] (https://pmc.ncbi.nlm.nih.gov/articles/PMC12391920/).* NCBI PMC*.
