Can AI Make Medical Errors? A Deep Dive into Accountability and Risk in Digital Health

Rasit Dinc

Rasit Dinc Digital Health & AI Research

Abstract

Introduction: The Promise and Peril of Artificial Intelligence in Medicine Artificial Intelligence (AI) is rapidly transforming healthcare, offering unp...

Introduction: The Promise and Peril of Artificial Intelligence in Medicine

Artificial Intelligence (AI) is rapidly transforming healthcare, offering unprecedented potential to enhance diagnostic accuracy, personalize treatment plans, and streamline administrative tasks. From analyzing medical images with superhuman speed to predicting patient outcomes, AI promises a future of reduced human error and improved efficiency. However, as AI systems become more integrated into clinical workflows, a critical question emerges: **Can AI make medical errors, and if so, who is accountable?** This question moves beyond mere technical failure to encompass profound ethical, legal, and regulatory challenges that the digital health ecosystem must address.

The Mechanisms of AI Error

Unlike human error, which often stems from fatigue, cognitive bias, or inexperience, AI errors typically arise from three primary sources:

1. **Data Bias and Quality:** AI models are only as good as the data they are trained on. If the training data is unrepresentative, incomplete, or contains historical biases (e.g., underrepresenting certain racial or gender groups), the resulting model will perpetuate and even amplify those biases, leading to inaccurate or inequitable outcomes for specific patient populations [1]. 2. **Model Opacity (The "Black Box"):** Many advanced AI models, particularly deep learning networks, operate as "black boxes." Their decision-making processes are complex and difficult for human clinicians to interpret or audit. When an error occurs, tracing the cause back to a specific input or internal logic can be nearly impossible, complicating error detection and correction [2]. 3. **Deployment and Integration Failures:** Errors can also occur when a well-trained model is deployed into a real-world clinical setting. Issues such

as poor user interface design, incorrect data input, or a mismatch between the model's training environment and the clinical environment can lead to misapplication and subsequent patient harm.

Ethical and Legal Accountability

The possibility of AI-induced medical errors forces a re-evaluation of traditional medical malpractice and accountability frameworks. In a conventional setting, the physician bears the primary responsibility. When an AI system is involved, the chain of accountability becomes complex, potentially involving:

The Developer/Manufacturer: *If the error is due to a flaw in the algorithm's design, biased training data, or inadequate validation, the software company may be liable under product liability law.* **The Clinician:** The physician who uses the AI tool is still responsible for the final clinical decision. They have a duty to exercise professional judgment and not blindly follow an AI recommendation. The concept of **non-maleficence** (do no harm) remains paramount. **The Hospital/Healthcare System:** *The institution is responsible for the proper procurement, integration, and oversight of AI tools, including ensuring staff are adequately trained and the system is functioning correctly.*

The challenge lies in determining the degree of autonomy granted to the AI. Is the AI merely a tool, or is it an autonomous decision-maker? Most current legal and ethical consensus views AI as a sophisticated tool, meaning the human clinician remains the ultimate decision-maker and is therefore primarily accountable [^3]. However, this consensus is continually tested as AI systems become more sophisticated. For instance, in cases where an AI system provides a diagnostic recommendation that a physician overlooks, and subsequent harm occurs, the legal analysis must weigh the physician's duty of care against the reliability and transparency of the AI tool. Furthermore, the concept of "shared liability" is gaining traction, where responsibility is distributed among the developer, the hospital, and the clinician, reflecting the multi-stakeholder nature of AI deployment in healthcare. This evolving legal landscape underscores the need for clear regulatory guidance that defines the expected standard of care when using AI.

The Regulatory Landscape and the Path Forward

Regulatory bodies worldwide are grappling with how to classify and govern AI in healthcare. The European Union's AI Act, for instance, classifies healthcare AI as "high-risk," imposing stringent requirements for data quality, transparency, human oversight, and robustness. The goal is to ensure that AI systems are safe, effective, and ethically sound before they reach the patient.

To mitigate the risk of AI-induced errors, a multi-pronged approach is necessary:

Mandatory Transparency: Developers must provide clear documentation on how models were trained, their limitations, and their performance metrics across diverse populations. **Human-in-the-Loop:** *Clinical workflows must retain a human oversight mechanism, ensuring that AI recommendations are*

reviewed and validated by a qualified professional. **Continuous Monitoring:** AI systems must be continuously monitored in real-world settings to detect performance drift or emerging biases that were not present during initial testing.

For more in-depth analysis on the intersection of digital health, AI governance, and the evolving regulatory frameworks, the resources at www.rasitdinc.com provide expert commentary and professional insights into navigating this complex landscape.

Conclusion

The answer to "Can AI make medical errors?" is unequivocally **yes**. AI systems, while powerful, are susceptible to errors rooted in data, design, and deployment. The true challenge is not preventing all errors—an impossible task for any technology—but establishing clear lines of ethical and legal accountability, ensuring robust regulatory oversight, and fostering a culture of critical human oversight. By embracing transparency and prioritizing patient safety, the healthcare industry can harness the transformative power of AI while responsibly managing its inherent risks.

**

[^1]: Chustecki, M. (2024). Benefits and Risks of AI in Health Care: Narrative Review. i-JMR*, 1(1), e53616. [^2]: Tighe, P. (2024). Artificial Intelligence and Patient Safety: Promise and Challenges.* PSNet*. [^3]: Pham, T. (2025). Ethical and legal considerations in healthcare AI.* Royal Society Open Science*, 12(1), 241873.