The Black Box of Blame: Can AI Be Held Accountable for Errors?

Rasit Dinc

Rasit Dinc Digital Health & AI Research

Abstract

Introduction: The Accountability Conundrum in the Age of AI The integration of Artificial Intelligence (AI) into professional sectors, particularly digi...

Introduction: The Accountability Conundrum in the Age of AI

The integration of Artificial Intelligence (AI) into professional sectors, particularly **digital health**, promises revolutionary advancements. However, this rapid adoption introduces a profound legal and ethical challenge: **who is responsible when AI makes a mistake?** The question of AI accountability is a critical hurdle for public trust. As AI systems become more autonomous and their decision-making processes more opaque—the "black box" problem—traditional liability frameworks are being stretched.

The Inadequacy of Traditional Legal Frameworks

Current legal systems primarily rely on three doctrines to assign fault: negligence, product liability, and vicarious liability. None of these perfectly fit the unique nature of AI errors [1].

1. Negligence and the Standard of Care

Negligence requires a breach of a duty of care, traditionally falling on the human professional. When an AI system provides faulty information, the physician who relies on it is often held solely responsible [2]. This is problematic: the physician may have acted within the standard of care, yet still be liable for an AI error due to flawed training data or algorithmic drift. The "learned intermediary" doctrine places an undue burden on the end-user as AI complexity increases [3].

2. Product Liability and the Evolving Defect

Product liability holds manufacturers responsible for defective products. While an AI algorithm is a "product," this framework is complicated by its capacity for continuous learning and self-modification. The algorithm that caused the

harm may not be identical to the one originally sold [4]. Proving a "defect" in a black-box neural network is difficult, shielding the developer from the scrutiny product liability laws are designed to impose.

3. Vicarious Liability and the Autonomous Agent

Vicarious liability, or *respondeat superior*, holds a supervisor responsible for a subordinate's actions. Treating AI as a subordinate, making the supervising institution or physician liable [5], struggles with highly autonomous AI. If an AI system makes an independent, unapproved decision that leads to harm, the human subordinate analogy breaks down, leading to a "responsibility gap" where no human or entity can be clearly blamed [6].

The Digital Health Imperative: Bias and Ethical Concerns

The accountability crisis is most urgent in digital health, where AI-driven tools are used for patient triage and treatment. A major ethical concern is **algorithmic bias**, where unrepresentative training data can cause errors that disproportionately affect certain demographic groups [7]. Accountability must extend beyond fault-finding to include an ethical duty of **transparency** and **explainability**. Professionals need to understand *why* an AI made a decision to address the root cause of an error—flawed data, poor design, or misuse.

For more in-depth analysis on this topic, the resources at www.rasitdinc.com provide expert commentary on the intersection of digital health, AI ethics, and emerging regulatory frameworks.

Emerging Solutions and the Path Forward

A new, dedicated framework is required to address the accountability gap. Several solutions are being debated, including **Shared/Common Enterprise Liability** to distribute fault across the entire value chain [8], **Mandatory Explainability (XAI)** to shift the burden of proof to the AI developer, and the creation of **AI-Specific Regulatory Bodies** for pre-market certification. A more radical proposal is granting AI a form of **Limited Legal Personhood** to directly address the "responsibility gap" [9]. The European Union's **AI Act** is a landmark attempt to address this by classifying AI systems based on risk, imposing strict requirements on high-risk applications like those in healthcare. This risk-based approach is a crucial step toward establishing clear, *ex ante* accountability, rather than relying on *ex post* litigation.

Conclusion: A Call for Proactive Governance

The answer to "Can AI be held accountable for errors?" is complex. Currently, no, AI itself cannot be held accountable in a meaningful legal sense. **Accountability is a human construct, and it must be assigned to the human or institution that controls the AI's design, deployment, or use.** The challenge is to move beyond fitting new technology into old legal boxes. A proactive, risk-based regulatory approach that mandates transparency and distributes liability across the value chain is essential to ensure the promise of AI in digital health is realized without sacrificing safety and justice.

References

[1] Defining medical liability when artificial intelligence is applied on diagnostic algorithms: a systematic review. [https://pmc.ncbi.nlm.nih.gov/articles/PMC10711067/] [2] Fault lines in health care AI – Part two: Who's responsible when AI gets it wrong. [https://carey.jhu.edu/articles/fault-lines-health-care-ai-part-two-whos-responsible-when-ai-gets-it-wrong] [3] Legal Handling of Artificial Intelligence in Radiology. [4] Legal Accountability in AI Failures and Malfunctions. [https://galkinlaw.com/legal-accountability-in-ai-failures/] [5] Artificial Intelligence in Healthcare: Legal and Ethical Considerations. [6] When AI fails, who do we blame? Attributing responsibility in human–AI interactions. [7] Health Equity and Ethical Considerations in Using Artificial Intelligence in Public Health. [https://www.cdc.gov/pcd/issues/2024/24_0245.htm] [8] A Common Enterprise Approach to AI Liability. [9] Artificial Intelligence and Medical Liability: A New Legal Personhood?