Navigating Liability in Automated Healthcare: Designing Multi-Stakeholder Framework for Responsibility in AI-Powered Health Care Systems

Ndage Kizito Nji, PhD

Department of Philosophy, Faculty of Arts, Letters, Social and Human Sciences (FALSH) The University of Yaoundé I

DOI: https://dx.doi.org/10.47772/IJRISS.2025.908000466

Received: 10 August 2025; Accepted: 16 August 2025; Published: 17 September 2025

ABSTRACT

Artificial intelligence (AI) is becoming a powerful tool in healthcare, offering faster diagnoses, better treatment planning, and improved patient care. But there is a serious ethical and legal problem: who is responsible if an AI system makes a mistake that harms a patient? Current legal systems were designed for situations where humans make decisions. They don't work well when responsibility is spread across many people and organizations, and the AI's decision-making is hidden inside a "black box." The old model, where the doctor is the only one held responsible, doesn't fit this new reality. This article argues that responsibility should be shared between three main groups — developers, healthcare providers, and regulators — under a principle called "technological due care". This means everyone involved must take active steps to prevent harm at every stage of the AI's life cycle. We will examine the weaknesses of current legal approaches, explain the ethical reasoning for a new system, and present a practical framework for safer AI use in healthcare.

INTRODUCTION

Over the past decade, artificial intelligence (AI) has moved from being a futuristic concept to an everyday reality in healthcare. Advances in machine learning, big data analytics, and computational power have allowed AI systems to take on increasingly complex tasks. These range from analyzing radiological images and predicting disease progression, to guiding robotic surgical procedures and generating personalized treatment plans tailored to a patient's genetic profile (Dilsizian & Siegel, 2021). The attraction is clear: AI can process vast amounts of data far faster than humans, detect subtle patterns that even experienced clinicians might overlook, and deliver recommendations in real time.

Advocates of medical AI paint a vision of a healthcare system that is more efficient, more precise, and more equitable. In this vision, diagnostic errors are dramatically reduced because AI can review thousands of past cases to identify the most likely diagnosis; treatment is personalized because algorithms can integrate a patient's genetic data, lifestyle factors, and medical history to recommend the most effective therapy; and healthcare resources are optimized because routine tasks, such as triaging patients or reviewing test results, can be handled automatically. AI could, in theory, free up physicians and nurses to focus on the human side of medicine — listening to patients, offering empathy, and building trust. In short, AI appears to promise a smarter, faster, fairer healthcare system.

However, the very features that make AI powerful also make it risky. Unlike a traditional medical device — say, a stethoscope or a pacemaker — AI systems are not static tools. They are adaptive: their performance changes as they are exposed to new data. This flexibility is often celebrated as a sign of progress, but it also introduces unpredictability. An algorithm that is accurate today may, after processing different kinds of patient data over months or years, begin to make systematically flawed recommendations — a phenomenon known as "model drift" (Pichler et al., 2021). Moreover, AI systems can fail in ways that are difficult to foresee. Consider a diagnostic algorithm designed to detect pneumonia from chest X-rays. If its training data came mostly from one demographic group, it might perform worse for patients outside that group — for example, producing more false negatives for women or for underrepresented ethnic populations (Char, 2018). These

failures are not like the mechanical breakdown of a device, where the fault is visible and traceable. They often emerge quietly, buried in complex data patterns, and may go unnoticed until they cause significant harm. The consequences of such failures are not hypothetical. In recent years, there have been documented cases where AI-based tools have misdiagnosed conditions, recommended unsafe treatments, or overlooked urgent medical needs. In one case, an AI system trained to predict which patients needed urgent care under-prioritized those with severe illnesses because it relied heavily on historical healthcare spending data — effectively disadvantaging low-income patients who had historically received less care, regardless of their actual health status (Obermeyer et al., 2019). These examples underline a stark truth: AI's potential to improve healthcare is matched by its potential to cause serious harm.

When AI does cause harm, the central legal and ethical question becomes: Who should be held responsible? In traditional medical practice, the answer is usually straightforward — the treating physician or surgeon is accountable for patient outcomes. The law assumes that the doctor exercised professional judgment based on available evidence, and if that judgment fell below the accepted standard of care, they may be found negligent. AI disrupts this assumption. In many modern AI systems — especially those powered by deep learning — the decision-making process is opaque even to the developers who built the system. This "black box" nature means that a doctor may not fully understand why an AI recommended a particular diagnosis or treatment. The doctor might trust the recommendation because the AI has an impressive track record — but if it turns out to be wrong, can the doctor be blamed for following it? Conversely, if the doctor ignores the AI's recommendation and harm occurs, can they be blamed for rejecting a tool that was statistically more accurate than humans in similar cases? This creates what some scholars call a "double bind" for clinicians (Cohen & Price, 2020). The challenge extends beyond doctors. Responsibility is potentially spread across a complex network of actors:

- The developer, who designed and trained the algorithm.
- The hospital or clinic, which purchased and implemented the AI system.
- The software company, which provided updates and technical support.
- The healthcare provider, who used the system in a clinical setting.
- Even the patient, who gave consent to its use in their care.

In this web of actors, pinpointing a single "responsible" party becomes not only difficult but sometimes impossible. This is what scholars refer to as the "responsibility gap" - a mismatch between our existing accountability frameworks and the realities of AI-driven decision-making.

Current liability models primarily medical malpractice and product liability are built on the assumption that there is a single, identifiable human or corporate actor whose actions caused the harm. In malpractice law, that actor is usually the physician; in product liability, it's the manufacturer of the defective product. But in AI healthcare, neither model fully captures the complexity of distributed, opaque, and evolving decision-making.

The problem is not merely legal but also philosophical. Traditional models of responsibility are rooted in an anthropocentric worldview: they assume that the decision-maker is a human being with moral agency - capable of intention, judgment, and foresight. AI systems, by contrast, have no moral awareness. They execute instructions, find statistical patterns, and generate outputs, but they do not "understand" or "intend" in the human sense. Holding an AI system itself responsible would therefore be both conceptually flawed and practically meaningless (Kohn, 2019). If the old frameworks do not fit, we must ask: What would a better one look like? One answer is to shift away from the search for a single culprit and towards a model of shared responsibility. In this model, liability and ethical obligations are distributed among all stakeholders, each according to their role in the AI system's lifecycle - from its design and training to its deployment and ongoing monitoring.

Central to this vision is the principle of "technological due care." This principle recognizes that preventing harm in AI healthcare requires constant vigilance, not only at the point of use but at every stage of the technology's life. Developers must ensure transparency and fairness in design; healthcare providers must maintain AI literacy and exercise independent judgment; regulators must enforce standards and monitor performance in the real world. Each group has a continuing duty to anticipate risks, communicate limitations,

and act swiftly when problems arise. This is a proactive approach. Rather than asking, after harm has occurred, "Who is to blame?", it asks before harm happens, "What steps are each of us taking to make sure this technology is safe and reliable?" In a system as high-stakes as healthcare, where errors can mean the difference between life and death, this shift from reactive blame to proactive care is not just desirable - it is essential.

This article takes the position that bridging the responsibility gap in AI healthcare requires more than small legal adjustments. It demands a rethinking of how we define responsibility itself in the age of intelligent machines. The discussion will proceed as follows: Section II will examine why existing legal models - malpractice and product liability - fail to adequately address AI-related harms. Section III will explore the philosophical and ethical reasons for moving towards a shared responsibility framework, including a discussion of why AI should not be granted legal personhood. Section IV will present the proposed multi-stakeholder liability model, detailing the specific duties of developers, healthcare providers, and regulators. Section V will discuss practical steps for implementing this framework, including legal reforms, training programs, and data-sharing mechanisms. Section VI will conclude with the broader implications for patient safety, public trust, and the future of automated medicine.

In short, the rise of AI in healthcare is not simply a technological shift - it is a moral and legal turning point. The choices we make now, in shaping the rules and responsibilities around AI, will determine not only how safely and effectively these systems are used but also whether the public trusts them at all. The task before us is clear: to build a framework that matches the complexity of AI with an equally sophisticated vision of accountability.

The Failure of Traditional Liability Models

When something goes wrong in medical care, lawyers traditionally use two main frameworks: medical malpractice and product liability. Both are built on the idea that harm can usually be traced to a human or a defective product. Yet when artificial intelligence (AI) is involved - especially in healthcare - these traditional systems often fail.

Medical Malpractice's Trouble with AI

Medical malpractice rests on the idea that a healthcare provider must meet a standard of care - that is, do what a reasonably careful professional would do in the same situation. If a doctor fails that standard and someone is harmed, the doctor may be held liable. But AI muddies this process. Clinicians often rely on AI tools that work like "black boxes"—even the people who created them cannot always explain how they reach conclusions. That makes it tough to know whether a doctor was negligent in following or rejecting the AI's recommendation. The expectation to act reasonably becomes unclear when the AI's logic is hidden (Smith, 2020).

A recent review of clinician experiences found that healthcare professionals felt ultimately responsible for patient outcomes but were uneasy about relying on obscure AI recommendations—particularly in rare or complex cases where lack of transparency can endanger care (Saoudi et al., 2025).

Furthermore, doctors risk becoming what some call "liability sinks." In complex systems that include automation, legal responsibility may fall unfairly on the human who interacts with AI—even when the fault lies elsewhere in the system (Sciolla et al., 2024). Clearly, traditional malpractice models aren't designed to deal with machines that think in ways people can't follow.

Product Liability Struggles with Evolving Systems

Product liability laws allow people harmed by faulty products to hold manufacturers responsible, even without proving intent. That approach works for defective pacemakers or surgical instruments-but AI isn't like those static tools. AI systems often change post-release. They learn from new data and adapt. This makes it hard to say a product was defective at the time it was sold. A system that behaves safely today might drift into error later—creating an unpredictable liability scenario (Sidley, 2024).

Moreover, product liability assumes there is a single maker to blame. AI systems in healthcare often involve developers, data providers, hospitals, and more. Blame quickly becomes diffuse, complicating responsibility

(Buchanan Ingersoll & Rooney, 2023). In the EU, even new directives try to address this. A 2024 update introduces a presumption that software defects causing harm may point to product defectiveness - but that still leaves grey areas in rapidly evolving AI systems (Health Law Sweden, 2025).

The Illusion of "Human-in-the-Loop"

Some argue that keeping a human in the loop—having a clinician approve or override AI decisions—solves the problem. The reasoning is that a human can catch errors before harm occurs. However, studies show that when humans think an AI system is reliable, they may trust it too much—and fail to question it—even when its suggestions make no sense. This phenomenon is called automation bias (Jama, 2023) . In a simulated pathology test, clinicians changed initially correct judgments after seeing incorrect AI advice—showing how automation bias can override human judgment (Rosbach et al., 2024). What's worse, vague assurances that "a human is always overseeing the AI" can give regulators and institutions a false sense of security. Carefully monitoring AI systems continuously is far safer than relying on a single human click, especially when trust in automation grows unchecked (Tendler et al., 2024) .

The Result: A Responsibility Gap

The failure of these models—malpractice, product liability, and human-in-the-loop—leaves a troubling void known as a responsibility gap. AI systems can cause harm, but current laws don't make responsibility clear: Was it the developer, the clinician, or the hospital? Or was it just a technical fluke beyond anyone's fault? This gap is harmful not just legally, but practically. A Stanford policy brief shows that fear of liability keeps physicians and organizations from adopting AI tools—even when those tools could improve care (Stanford HAI, 2024). That means patients might be denied benefits due to legal uncertainty. Consider a real example: Google's healthcare AI model "Med-Gemini" misdiagnosed a non-existent brain structure, "basilar ganglia." The mistake slipped through unnoticed—even study authors missed it—and Google quietly corrected the blog but not the official paper. That kind of "hallucination" can propagate harmful cascades if clinicians don't catch them—but who is liable when it happens? (The Verge, 2025). This combination of elements—opaque AI, shifting designers, automation bias, and legal uncertainty—means that harm may occur, but accountability remains elusive. Patients may be harmed, but no one is clearly responsible. That undermines trust and fails to protect healthcare consumers.

Traditional liability models were built for a world where humans made decisions directly and products were static and traceable. AI in healthcare changes that world. AI is adaptive, opaque, and often enhances but shifts human judgment rather than replacing it. Medical malpractice fails when we can't judge whether clinicians acted reasonably because we don't understand how AI forms its opinions. Product liability fails when AI changes or evolves after release—making it unclear if the product was ever defective. Human-in-the-loop fails when automation biases cloud clinician judgment or reduce oversight effectiveness. Together, these weaknesses produce a legal blind spot—a responsibility gap—where harm can happen with no clear path to accountability or compensation. To protect patients and enable safe AI innovation in healthcare, we must build new legal frameworks suited to the complexity of AI.

The Philosophical and Ethical Foundations of a New Framework

When AI is used in healthcare, we face challenging questions about responsibility that our traditional ways of thinking can't address. In the past, responsibility meant pointing to a doctor or a manufacturer when something went wrong. But AI upends those familiar patterns by being complex, opaque, and adaptive. This means we can't rely on assigning blame to a single individual or entity anymore.

One tempting but flawed idea is to treat advanced AI as if it were a legal person—capable of being held accountable on its own. Some argue that this could simplify liability issues. But AI systems lack consciousness, intent, or moral thinking. They can't understand right or wrong, learn from remorse, or make amends. Giving them legal personhood risks letting humans dodge responsibility for harmful outcomes (Oxford Academic, 2025). Scholars argue convincingly that this view is ethically and practically unsound (Wikipedia Ethics_of_AI).

Because AI cannot bear moral weight itself, responsibility must stay with humans and institutions. But AI systems are often the product of many people: developers, data scientists, hospitals, and regulators all play a role. This distributive nature of responsibility calls for a shift in how we think about liability—away from blame assigned to a single person, and toward shared responsibility across stakeholders.

A key ethical concept here is what we might call technological due care—a shared duty throughout the life of an AI system to prevent harm before it happens. This means everyone involved must act proactively, sensitively, and ethically at every stage of AI design, deployment, and oversight. Developers must ensure fairness and traceability in their systems; providers need to understand the AI's strengths and limitations; and regulators must create standards for safety, transparency, and equity (BMC Medical Ethics, 2025) .

This approach also responds to the so-called "AI-chasm," where a system may perform well in controlled trials but fail in real-world settings due to unanticipated variables. A proper ethics framework can help bridge that gap through pilot testing, phased deployment, and interdisciplinary oversight at every stage of implementation (AI & Society, 2025).

Another important challenge is the diffusion of responsibility. Because AI outcomes emerge from complex systems with multiple contributors, traditional models of assigning blame do not work well (AI and Ethics, 2022). Political philosopher Barman and collaborators describe this as a "responsibility gap" in AI-driven healthcare—one that arises from opaque systems, extended causal chains, and uncertainty in decision-making (Journal of Medical Ethics, 2025).

But acknowledging complexity alone is not enough. We need ethical clarity. One helpful idea is to map responsibilities clearly: developers are responsible for transparency and fairness in design; providers are responsible for oversight and patient communication; regulators are responsible for setting standards and enforcing them; and institutions are responsible for implementing governance structures. This helps prevent diffused accountability—where everyone is partly responsible but nobody is accountable (AI & Society, 2022).

Ethics also demands that AI systems in healthcare preserve human dignity and care. AI should not replace the relationship between doctors and patients. Instead, it must support human values like empathy, trust, and respect (Forbes, 2024). AI must be designed to enhance—not erode—the human-centered essence of medicine. This requires us to embed ethical concerns, such as privacy, equity, and trust, into every stage of AI design and use (CDC, 2024).

A world where AI decision-making remains partially mysterious requires transparency and traceability. If we can't follow how AI makes its decisions, we lose trust—and accountability. Ethical frameworks insist that providers can access meaningful explanations, and systems are designed with audit trails and clarity (World Economic Forum, 2025).

Finally, the ethical framework must avoid abstractions and translate into real actions. The "human warranty" approach puts human oversight at the center, but that alone doesn't guarantee equitable outcomes unless it is operationalized through clear protocols and training (Springer article, 2023).

In sum, the philosophical and ethical foundation for a new approach to liability in healthcare AI rests on four central ideas:

- Reject legal personhood for AI—responsibility always lies with humans and institutions.
- Embrace shared, distributed responsibility—everyone in the system has duties.
- Anchor ethics in technological due care—ethical foresight built into every phase of AI life.
- Preserve human-centered values, fairness, transparency, and accountability in practice.

These principles set the stage for the multi-stakeholder liability model to come in the next section—one that aligns responsibility with complex system structures while ensuring patient safety and institutional integrity.

A Multi-Stakeholder Model of Liability,

When AI causes harm in healthcare, the answer isn't to find a single culprit—but to recognize that responsibility must be shared. A multi-stakeholder model of liability spreads accountability across developers, healthcare providers, and regulators. This reflects how AI systems are built, used, and governed—they are not the product of a single actor, but the outcome of many roles working together or failing together (Springer, 2024).

First, developers and vendors must take a proactive role in safety, transparency, and fairness. This isn't about assigning blame only when something breaks; it means building systems with high-quality, bias-free datasets, clear documentation, and mechanisms for identifying and fixing downstream issues. Some contracts now reflect this in shared liability clauses, where vendors accept risk for design flaws and hospitals take responsibility for misuse or bypassing system guidance (AFS Law, 2025). Nor should AI systems be "trust me products"—developers must maintain ongoing oversight, perform audits, and update systems to address flaws post-deployment (Stanford Health Policy, 2025).

Second, healthcare providers (clinicians and hospitals) must exercise informed, responsible oversight. They can't treat AI as infallible. Human judgment remains essential in interpreting and contextualizing AI recommendations. Oversight protocols—such as sign-offs, peer review, or multidisciplinary governance—help provide checks and balances. Organisations like hospitals need formal structures (e.g., ethics boards, AI review committees) to evaluate AI tools, implement policies, and train staff (Ethics & Governance of LLMs, WHO, 2025). Without these, human-in-the-loop becomes a weak illusion, and accountability suffers.

Third, regulators and policy makers serve as the guardians of public safety. They must move beyond approving AI systems at a single point in time. Instead, they need multi-stakeholder, proactive governance frameworks that keep pace with AI advances and evolving risks. Structured approaches—like algorithm assurance labs—can test AI tools against standard criteria such as fairness, robustness, and safety before widespread use (JAMA, 2024). Regulatory regimes should evolve beyond traditional device approval to include audits, performance tracking, and real-world evidence requirements (Stanford HAI Brief, 2024).

The strength of this multi-stakeholder model lies in its interlocking safeguards. If a developer fails to account for bias, providers may catch it via clinical oversight. If providers misapply the AI, regulators can hold them accountable. If regulators fail, praxis in clinical settings—through malpractice insurance or internal review—can provide compensation or remediation. The result is liability overlaps rather than gaps, reducing the chance that any single failure leads to uncompensated patient harm (Springer, 2024).

Importantly, shared liability must not lead to "diffusion of responsibility," where everyone assumes someone else is on the hook. Clear role definitions help prevent that. For instance, contracts may stipulate that providers are responsible for misuse, vendors for errors in the algorithm beyond intended use, and regulators for oversight gaps (AFS Law, 2025). Legal frameworks can support this by enforcing responsibilities based on capacity and influence—for example, holding parties partially liable unless they can show they met their agreed standards.

Consumer health AI offers a promising field to test shared governance models. A recent expert consensus process led to the proposal of a Health AI Consumer Consortium (HAIC²)—a multi-stakeholder body combining developers, providers, regulators, and patient advocates to guide policy, accountability, and patient-centered governance (Yesil Science, 2024). Such collaborative models align stakeholder incentives with public safety and trust.

One key philosophical principle underlying this model is that of fiduciary duty. Each stakeholder largely benefits from AI's value—developers from innovation, providers from efficiency, regulators from modernization—so each must also bear a corresponding duty to prevent harm. This concept echoes proposals to adapt contract or tort law to require shared accountability in complex tech systems, unless proven otherwise (Springer, 2024).

Moreover, frameworks like FUTURE-AI emphasize the ethical foundation of this model. FUTURE-AI outlines six guiding principles—fairness, universality, traceability, usability, robustness, and explainability—to be embedded throughout the AI lifecycle from design to deployment. These principles help distribute moral and practical duties across stakeholders (ArXiv, 2023).

We must also recognize the moral outsourcing problem—the tendency to blame AI systems themselves for ethical failings. This deflects focus from the humans who built, approved, or deployed the systems. A multi-stakeholder framework confronts this by insisting on human accountability at every step (Wikipedia, Moral Outsourcing).

In practice, this model also benefits innovation. When liability is shared fairly, developers and providers aren't discouraged from exploring new tools. They can move forward with ethically grounded confidence, knowing responsibilities are structured, understood, and aligned with patient welfare—not deferred or hidden.

- To summarize, a multi-stakeholder liability model for AI in healthcare requires:
- Developers ensuring safety, transparency, updates, and shared risk.
- Providers exercising caution, context-aware judgment, and placing patients' welfare first.
- Regulators providing ongoing, evidence-based oversight and public safeguards.

Together, these overlapping layers of responsibility aim to protect patients, encourage thoughtful innovation, and maintain public trust in AI-assisted healthcare.

Practical Implementation and Policy Implications

Turning the idea of a multi-stakeholder model of liability into reality requires careful planning and strong commitment from all actors in the healthcare and technology sectors. It is not enough to say that developers, healthcare providers, and regulators must share responsibility; there must also be clear processes, rules, and systems that make this sharing of responsibility possible in practice. Without such systems, the principles will remain only on paper, and patients will still face harm without adequate remedies.

One of the first steps in practical implementation is creating clear laws and regulations that define the duties of each stakeholder. At present, many countries have fragmented legal rules for healthcare liability, technology use, and medical devices, but these do not fully address AI-specific challenges (Floridi et al., 2018). Legislators must establish rules that clarify who is responsible when an AI system makes a mistake, how evidence can be collected from AI "black boxes," and how patients can claim compensation. This means adopting laws that reflect the concept of technological due care, which requires all actors to act with a level of caution appropriate to the complexity and risks of AI systems (Calo, 2015).

Informed consent is another critical area. Patients must know when AI is involved in their diagnosis or treatment and must have the option to ask questions or refuse AI assistance if they feel uncomfortable (Beil et al., 2019). This requires providers to explain in plain language what the AI system does, its potential benefits, and its limitations. Consent forms may need to include specific clauses on AI usage, much like current forms include clauses about surgical risks or anesthesia. Without informed consent, the use of AI in healthcare risks violating patient autonomy and trust.

Another practical measure is improving transparency and data sharing among stakeholders. Developers often treat AI algorithms as trade secrets, making it difficult for healthcare providers or regulators to understand how the systems work or detect potential flaws (Price et al., 2019). To address this, policymakers can require "algorithmic impact reports" and periodic audits. These reports could detail the training data used, the performance of the AI system across different patient groups, and any updates made to the algorithm. Data sharing could also extend to creating anonymized patient datasets for independent verification, ensuring that AI tools work safely across diverse populations.

For situations where responsibility is shared or unclear, no-fault compensation systems can be valuable. In such systems, patients harmed by AI-assisted care receive compensation without having to prove negligence

from a specific party (Brownsword & Goodwin, 2012). This avoids lengthy court battles and ensures that victims are not left without help. Funding for these systems could come from a pooled contribution by developers, healthcare institutions, and government agencies, similar to how vaccine injury compensation funds operate in some countries.

International cooperation is also essential. Many AI healthcare systems are developed in one country, deployed in another, and updated through cloud-based services that operate globally (European Commission, 2021). Without cross-border standards, there is a risk of inconsistent safety levels and legal uncertainty. International bodies such as the World Health Organization could help establish global guidelines on AI safety, testing, and liability. These guidelines could be voluntary at first but later integrated into national laws through treaties or harmonization agreements.

From a policy perspective, the multi-stakeholder model also requires capacity building. Healthcare providers must be trained in AI literacy so they understand how to interpret AI recommendations and when to override them (Gerke et al., 2020). Developers must be trained in ethical design principles and the social impact of their products. Regulators need expertise in both medical law and computer science to effectively oversee AI tools. This training could be built into medical school curricula, continuing professional development programs, and certification processes for developers.

Finally, public engagement is critical. If AI is to be trusted in healthcare, patients and citizens must feel that their voices are heard in shaping the rules. This can be achieved through public consultations, patient advocacy groups, and citizen panels that advise on AI governance. Public input can highlight cultural, ethical, and local concerns that might not be visible to policymakers or developers working in corporate or academic settings.

In sum, the practical implementation of a multi-stakeholder liability model for AI in healthcare requires a blend of legal reform, ethical safeguards, technical transparency, patient empowerment, international collaboration, professional training, and public engagement. The policy implications are far-reaching: shifting from a blame-focused system to a shared responsibility model will require political will, investment in infrastructure, and a willingness to rethink traditional notions of medical accountability. But with these steps, it is possible to ensure that AI in healthcare fulfills its promise while minimizing its risks. As Bryson and Winfield (2017) note, the goal is not only to assign blame after harm occurs but to design systems where harm is far less likely in the first place.

CONCLUSION

The integration of AI into healthcare brings both extraordinary opportunities and serious responsibilities. On one hand, AI tools can improve diagnosis, predict patient outcomes, and make healthcare more efficient than ever before. On the other hand, they also raise complex questions about safety, fairness, and accountability (HIMSS, 2024). The multi-stakeholder governance model offers a balanced path forward. By involving developers, clinicians, patients, regulators, and institutions in joint decision-making, we can ensure AI is not just technically sound but also ethically aligned. This approach acknowledges that no single actor can fully guarantee safe and fair AI; responsibility must be shared in a coordinated and transparent way (Portulans Institute, 2024). Key principles such as transparency, patient consent, equity, and continual oversight must become embedded in everyday practice. Informed consent should reflect AI's role in patient care so people understand how decisions are being made about them (CAPP Physicians, 2024). Regular monitoring and certification processes can help detect problems early, avoiding harm while maintaining trust (JAMA, 2025). This ethical vision does not mean slowing down innovation. On the contrary, it means creating the guardrails that allow safe, meaningful progress. When developers build transparency into their tools, when hospitals apply governance frameworks, and when regulators set adaptive standards, AI can grow in ways that improve health outcomes for all (Forbes, 2025). The road ahead will require international cooperation. AI is not confined by national borders, and healthcare is increasingly global. Shared standards, cross-border data governance, and collaborative ethical reviews will help ensure that AI tools are both safe and effective wherever they are deployed (FUTURE-AI, 2023). In the end, the promise of AI in healthcare will only be realized if trust is built and maintained. That trust grows when every stakeholder—patients, doctors, engineers, policymakers—plays their part. Governance is not just a policy tool; it is the bridge between technical

innovation and ethical care. If we can achieve that synergy, AI will not replace the human side of healthcare. Instead, it will strengthen it—helping us deliver care that is more accurate, fair, and compassionate than ever before.

REFERENCES

1. Adadi, A., & Berrada, M. (2018). Peeking inside the black-box: A survey on Explainable Artificial Intelligence (XAI). IEEE Access, 6, 52138-52160.
2. AFS Law. (2025). AI Service Agreements in Healthcare: Indemnification Clauses Explained.
3. Beil, M., Proft, I., van Heerden, D., Sviri, S., & Markota, M. (2019). Ethical considerations about artificial intelligence for prognostication in intensive care. Intensive Care Medicine Experimental, 7(1), 70.
4. Brownsword, R., & Goodwin, M. (2012). Law and the Technologies of the Twenty-First Century. Cambridge University Press.
5. Bryson, J., & Winfield, A. (2017). Standardizing Ethical Design for Artificial Intelligence and Autonomous Systems. Computer, 50(5), 116–119.
6. Calo, R. (2015). Robotics and the Lessons of Cyberlaw. California Law Review, 103(3), 513–563.
7. Calo, R. (2017). A New Liability Regime for AI-Driven Products. In The Law and Ethics of AI and Robotics (pp. 23-45). University of Cambridge Press.
8. Char, D. (2018). Algorithmic Bias in Healthcare. JAMA, 319(24), 2568-2569.
9. Cohen, I. G., & Price, W. N. (2020). Black box medicine: The legal and ethical implications of AI in medical diagnosis. American Journal of Law & Medicine, 46(2-3), 237-268.
10. Dilsizian, S. E., & Siegel, E. L. (2021). The Future of Radiology: Artificial Intelligence and the Radiologist. Radiology, 298(3), 515-516.
11. European Commission. (2021). Proposal for a Regulation laying down harmonised rules on artificial intelligence (Artificial Intelligence Act). Brussels.
12. Fenech, B., et al. (2018). AI in Medicine: A Call for a Re-evaluation of Informed Consent. Journal of Medical Ethics, 44(10), 653-659.
13. Floridi, L., Cowls, J., Beltrametti, M., et al. (2018). AI4People—An Ethical Framework for a Good AI Society. Minds and Machines, 28(4), 689–707.
14. Forbes Tech Council. (2025). Building Trust in AI in Healthcare: The Critical Role of Responsible Adoption.
15. FUTURE-AI Consortium. (2023). Six Guiding Principles for Responsible AI. arXiv preprint.
16. Gasser, U., & Senden, M. (2019). AI-Based Decision-Making in Medicine: A Policy and Regulatory Framework. Harvard Journal of Law & Technology, 32(2), 527-584.
17. Gerke, S., Minssen, T., & Cohen, I. (2020). Ethical and legal challenges of artificial intelligence-driven healthcare. Artificial Intelligence in Healthcare, 295–336.
18. JAMA Editorial. (2024). Legal Risks and Rewards of Artificial Intelligence in Healthcare. JAMA.
19. Kessler, D. P., & McClellan, M. (2002). The Effects of Malpractice Liability on the Delivery of Medical Care. Journal of the American Medical Association, 288(16), 2098-2101.
20. Kohn, A. (2019). Should AI Be Held Liable for its Actions? A Critical Analysis. Journal of International Affairs, 72(2), 1-15.
21. Pichler, M., et al. (2021). Algorithmic Bias in Clinical Prediction Models. Journal of the American Medical Informatics Association, 28(11), 2390-2397.
22. Portulans Institute. (2024). Multi-Stakeholder AI Governance Models.
23. Price, W. N., Gerke, S., & Cohen, I. G. (2019). Potential liability for physicians using artificial intelligence. JAMA, 322(18), 1765–1766.
24. Stanford Health Policy Brief. (2025). Understanding Liability Risk in Healthcare AI. Stanford University.
25. Sutton, A. (2021). De-Skilling the Doctor: The Unintended Consequences of AI in Medicine. The New England Journal of Medicine, 384(15), 1401-1403.
26. WHO Report. (2025). Ethics and Governance of AI for Health.