USOD PUBLISHERS
United Scientific Open Discovery

Journal of Surgical Outcomes and Innovation
Volume 2 | Issue 1

Review Article

First to Market, First to Fail: Artificial Intelligence in Healthcare and the Erosion of Clinical Accountability

Andrew J Fishman[1,2]

[1]Department of Otolaryngology Head & Neck Surgery, University of Missouri Medical Center, USA

[2]Department of Engineering, University of Johannesburg, South Africa

***Corresponding author:** Andrew J Fishman, Department of Otolaryngology Head & Neck Surgery, University of Missouri Medical Center, USA.

Citation: Andrew J Fishman (2026) First to Market, First to Fail: Artificial Intelligence in Healthcare and the Erosion of Clinical Accountability. J of Sur Out & Inno 2(1), 1-7.

Received Date: March 29, 2026
Accepted Date: March 31, 2026
Published Date: April 10, 2026

Keywords: Artificial Intelligence, Clinical Liability, Medical Malpractice, Algorithmic Accountability, Informed Consent, Standard of Care, Black-Box Medicine, Healthcare Law

Introduction

Artificial intelligence (AI) in healthcare is no longer speculative—it is accelerating rapidly and entering the clinical mainstream. From diagnostic tools to predictive algorithms, AI systems are being embedded into workflows once governed solely by human judgment. This shift promises increased accuracy and efficiency, but it also invites new risks that challenge the assumptions at the heart of medical law, ethics, and professional responsibility.

As legal scholar Frank Pasquale warns, we are increasingly governed by algorithmic systems whose inner workings remain opaque even as they wield enormous influence. 'Important corporate actors have unprecedented knowledge of the minutiae of our daily lives, while we know little to nothing about how they use this knowledge to influence the important decisions that we—and they—make' [1]. This asymmetry erodes the traditional expectations of informed oversight and accountability in medicine. As Pasquale further observes, transactions that become too complex to explain to outsiders may well be too complex to be permitted to exist at all [1]. In the context of healthcare, where trust and transparency are foundational, the rise of inscrutable decision-making systems threatens to displace the physician's role as an interpreter of risk with that of an operator within a black box. Pasquale captures the essential danger: the contemporary information economy 'more closely resembles a one-way mirror' in which powerful institutional actors accumulate knowledge about individuals while remaining opaque themselves [1].

These concerns are not abstract. History offers concrete warnings about what happens when technological enthusiasm outpaces empirical validation. High-profile failures like Theranos and IBM's Watson for Oncology demonstrate how premature deployment, exaggerated marketing claims, and lack of independent oversight can lead to real-world harm. These failures are not outliers. They reflect a recurring pattern in biomedical innovation: the rush to be first to market—driven by competition for funding, prestige, or regulatory positioning—can incentivize ethical shortcuts in pursuit of rapid validation [2,3].

High-Profile Failures: Theranos and IBM Watson for Oncology

Theranos promised to revolutionize blood testing with a device that could run hundreds of diagnostic tests from a single drop of blood. Its founder, Elizabeth Holmes, raised hundreds of millions in venture capital and forged high-profile partnerships with institutions including Walgreens, Safeway, and the U.S. Department of Defense. Retired General James Mattis, a former Secretary of Defense, joined the company's board and publicly supported its potential military applications. However, the technology was never validated, and the devices were never used in clinical or combat settings. Mattis later testified at Holmes's trial that he had been misled about the system's capabilities [4].

Theranos claimed it could perform a wide range of tests using its proprietary Edison device. In reality, most tests were conducted on traditional lab machines, and the Edison device proved unreliable and inaccurate. When independent investigations revealed the scope of the deception, the fallout was swift. Patients were misdiagnosed, test results were fabricated, and the company collapsed under criminal charges and public scrutiny (U.S. Securities and Exchange Commission) [5,6]. The catalyst was a groundbreaking 2015 Wall Street Journal exposé, which revealed that 'Theranos has struggled behind the scenes to make its technology work and has routinely used machines made by other companies for most of its testing' [7]. Holmes was ultimately convicted in January 2022 on multiple counts of fraud (Carreyrou, 2018; U.S. Securities and Exchange Commission) [4,6].

Similarly, IBM's Watson for Oncology was promoted as a powerful AI clinical decision support tool intended to help physicians diagnose cancers and recommend treatment protocols. However, internal IBM documents later revealed that Watson produced 'unsafe and incorrect' cancer treatment suggestions, including advice that conflicted with accepted medical guidelines [8]. These documents further show that many recommendations stemmed from training on synthetic or curated hypothetical cases rather than broad, real-world patient datasets [8]. Physician-users and hospital partners reported inconsistencies, flagged errors, and questioned the tool's reliability in routine clinical settings [9].

From a legal and regulatory perspective, Watson's design and deployment exposed significant liability concerns. Scholars note that existing tort doctrines struggle to manage the opacity

of black-box AI in healthcare—particularly when establishing causation, identifying responsible parties, or proving negligence [10,11]. Reviewers have also proposed adapting product liability standards to account for AI design flaws that may directly harm patients [12,13].

At MD Anderson Cancer Center, the collaboration with IBM reportedly cost over $60 million before being quietly terminated due to poor performance and technical integration issues [14]. Memorial Sloan Kettering Cancer Center played a key role in training Watson, and critics argued that the AI's recommendations reflected MSK's institutional preferences rather than generalized evidence-based consensus [3]. Jupiter Medical Center in Florida also adopted Watson early on, but later scaled back its use after clinicians found the AI's advice inconsistent with standard practice [3].

The episode illustrates how institutional prestige can lend credibility to immature technologies, and how the drive to innovate can sometimes outpace the due diligence needed to ensure patient safety [5,8]. Even OpenAI CEO Sam Altman, in his May 2023 testimony before the U.S. Senate, acknowledged the magnitude of these risks: 'I think if this technology goes wrong, it can go quite wrong, and we want to be vocal about that' [15].

As ethicist Wendell Wallach has observed, the pace of technological development persistently outstrips our capacity to develop adequate ethical and legal oversight—a pattern that biomedical innovation repeats with troubling regularity [16]. These cases remind us that bold claims in health technology must be backed by transparency, peer review, and a clear understanding of the consequences when those systems fail.

The Paradox of Human Oversight in Clinical AI

At the center of these challenges lies a paradox. Regulators and ethicists often invoke the principle of a 'human in the loop' as a safeguard, assuming that a physician will oversee AI-generated recommendations, exercise independent judgment, and act as a moral buffer between machine logic and human care. But this framework collapses under scrutiny. AI systems now process data at scales and speeds far beyond human comprehension. Their logic is often probabilistic, statistical, or buried in neural networks that are inaccessible even to their creators [17]. In practice, the human may remain 'in the loop' nominally, but not meaningfully.

This paradox is not merely theoretical. It is embedded in law. Under the European Union Artificial Intelligence Act, high-risk AI systems must be designed to ensure that 'natural persons can oversee the functioning of the system,' including the ability to 'interpret the system's output,' 'intervene,' and 'decide not to use the system or override or reverse the output' [18]. This mandate reflects a vision of oversight rooted in an earlier technological paradigm—one in which systems were predictable, deterministic, and fully within the grasp of human reasoning. As AI systems increasingly rely on deep learning and operate at machine speeds, the legal requirement of oversight becomes practically impossible to fulfill in real-time clinical contexts.

This paradox collides with long-standing legal doctrines. Chief among them is the learned intermediary doctrine, which holds that a manufacturer's duty to warn of medical risks is satisfied by informing the physician, who acts as a knowledgeable gatekeeper. Yet this doctrine emerged in an era when diagnostic and therapeutic tools were static, interpretable, and clearly under human control. Adaptive AI systems—particularly those based on black-box models—undermine these foundational assumptions. If the physician cannot meaningfully interrogate, validate, or override the AI's output at the point of care, then invoking them as a legal intermediary becomes a form of legal fiction [19,20]. The law assumes that the physician remains in control. The technology reveals that this control may be illusory.

This article explores that fiction and the real challenges it obscures. Drawing on doctrinal analysis, real-world case studies, and cross-jurisdictional regulatory frameworks, we examine the evolving terrain of liability in AI-driven healthcare. Our aim is not to halt innovation, but to ask whether our legal architectures can absorb it without distortion. As OpenAI CEO Sam Altman stated in his written testimony before the U.S. Senate in May 2023: 'We believe it is essential to develop regulations that incentivize AI safety while ensuring that people are able to access the technology's many benefits' [15].

Legal Foundations of Clinical AI Liability

The cautionary lessons of Theranos and IBM's Watson for Oncology show how unchecked technological enthusiasm can morph into disaster. Both stories expose how hype, prestige, and competitive pressure can propel unready technologies into clinical spaces long before their safety and reliability are established.

Tort law has long been the backbone of civil liability in medicine, guiding how courts determine responsibility for harm. Its core strength has been clarity: the ability to trace an injury back to a human act or to a static product defect. The rise of artificial intelligence unsettles that structure. AI is not static but dynamic, not transparent but opaque, and often co-constituted by both human and machine agency. These features challenge the legal system's assumptions about traceability and control, raising profound questions about how liability should be assigned when outcomes emerge from algorithms that adapt, evolve, and influence clinical judgment in ways no traditional tool ever has.

Medical Malpractice and the Question of AI-Augmented Judgment

Medical malpractice law has long rested on a stark premise: when a physician's conduct falls below the accepted standard of care and harm follows, liability attaches. Courts traditionally ask whether a reasonably prudent physician, under similar circumstances, would have acted differently—an inquiry guided by expert testimony, clinical guidelines, and the core assumption that physicians are autonomous agents making informed, traceable choices [21,22].

Artificial intelligence destabilizes this foundation. Clinicians now work in an environment where decisions may be co-authored with black-box systems whose reasoning is invisible, whose training data may be skewed, and whose recommendations may diverge from accepted clinical standards. When such systems prove flawed—as history shows they can—the negligence framework collapses into paradox. Is a physician negligent for following a defective AI output, or for disregarding one that hindsight shows would have prevented harm?

Physicians may soon face claims that not using AI constitutes negligence, even as the use of AI itself exposes them to risks when the system's errors are revealed. The profession is thus being pressed into a legal crucible: courts must decide whether the 'reasonable physician' is one who trusts the algorithm or one who resists it. What was once a doctrine built on clear human agency now confronts a world where responsibility is diffuse, and where the clinician is forced into partnership with a machine that may be brilliant, biased, or catastrophically wrong.

Product Liability and Defective Algorithms

Where malpractice focuses on human conduct, product liability governs defects in the tools and devices used in care. Traditionally, three categories of defect are recognized under the Restatement (Third) of Torts: design defects, manufacturing defects, and failure to warn [23].

- Design defects may occur when the structure or logic of an AI model—such as its training data, labeling assumptions, or optimization parameters—leads to systematically harmful outputs, even if the system behaves as intended.
- Manufacturing defects raise more elusive questions for software. Updates, corrupted training inputs, or unintended system interactions may be analogized to physical flaws, but the evolving nature of AI complicates these analogies.
- Failure to warn is especially fraught. The duty to inform users of non-obvious risks assumes that the developer understands the system well enough to issue such warnings. Yet many high-performing AI models—particularly black-box neural nets—defy straightforward explanation, even by their creators. As Wagner observes, emerging technologies expose gaps in conventional liability doctrine by introducing tools whose risks are opaque even to their manufacturers [24].

The case of Taylor v. Intuitive Surgical, illustrates how courts may adapt these principles in complex technological settings [25]. The da Vinci surgical robot, while not an AI-enabled system, represented one of the first widely adopted examples of robotic surgery. The Washington Supreme Court held that manufacturers owe a duty to warn purchasing hospitals in addition to operating physicians, reasoning that hospitals must be informed of the risks of complex devices in order to credential physicians appropriately and provide optimal patient care. The lesson for AI is clear: as systems grow more opaque and influential, liability frameworks will likely demand shared responsibility across the entire enterprise [13,23,24].

Dual Exposure and the Search for Legal Reform

The cautionary lessons of Theranos and Watson for Oncology showed how enthusiasm and institutional prestige can blur the lines of accountability, leaving patients to suffer while responsibility diffuses into the shadows. AI threatens to magnify this problem. A flawed diagnosis or treatment recommendation might involve a physician who trusted the output, a developer whose training data or model design introduced bias, and a hospital that embedded the system into its clinical pathways without real-time auditing.

This shared exposure has prompted a range of reform proposals:

- Enterprise liability: Rather than isolating physicians, responsibility could shift to the organizations that profit from or deploy AI systems. This approach recognizes that institutions shape how AI is integrated into practice and are often better positioned to absorb liability costs [26].
- No-fault compensation schemes: Modeled on the U.S. vaccine injury program, these frameworks could ensure swift recovery for harmed patients while minimizing litigation burdens. By decoupling compensation from fault, they may also reduce the chilling effect on innovation [27].
- Dynamic monitoring obligations: Developers and institutions may be tasked with continuous surveillance of deployed systems, ensuring that adaptive models are tracked for new risks and that post-market harms are promptly addressed.

Together, these proposals reflect a stark truth: clinical AI does not simply add new tools to medicine, it alters the very architecture of responsibility. The legal system faces a choice between clinging to doctrines built for static technologies or embracing reforms that acknowledge the speed, opacity, and distributed nature of AI.

Opacity, Black-Box Systems, and the Collapse of Legal Accountability

The integration of AI into healthcare reconfigures the evidentiary logic that underpins the entire system of liability. Traditional malpractice and product claims rely on the ability to reconstruct what happened and why. But many AI systems, especially those based on deep learning, offer no clear path to understanding their outputs. This creates a dangerous asymmetry: clinicians and patients are expected to trust systems whose reasoning is inaccessible, even in the aftermath of harm.

These so-called 'black box' models—trained on vast datasets with millions of parameters—may outperform human diagnosticians in pattern recognition tasks, but they do so without generating interpretable rationales. This lack of transparency undermines the legal imperative of accountability. In litigation, both plaintiffs and defendants are entitled to evidence that explains causation. If an AI system misguides care but cannot be deconstructed in court, how can responsibility be assessed?

In practice, opacity hinders multiple legal processes. Discovery suffers because plaintiffs may be unable to subpoena usable records of how the AI reached its conclusion. Expert testimony becomes compromised because experts testifying about standard of care may be unable to explain or evaluate the AI's conduct. Comparative fault becomes nearly impossible when neither the clinician nor the developer can explain how or why an AI output was generated.

Opacity also disrupts informed consent. Patients may be unaware that an AI tool played a role in their diagnosis or treatment, and even if disclosure is made, the nature of that influence may be beyond their comprehension. This raises an ethical and legal question that cuts to the heart of medical practice: can a patient truly give informed consent to an invisible co-decision-maker?

As Burrell argues, opacity in machine learning is not just a technical flaw—it is a structural feature of modern systems [28]. Wachter, Mittelstadt, and Floridi similarly warn that transparency and accountability are prerequisites for trustworthy AI, yet remain elusive in practice [17]. These epistemic gaps make it more difficult to satisfy tort's traditional standards of foreseeability, causation, and fault.

The courtroom dimension of this crisis is severe. Traditional malpractice litigation relies on an orderly progression: duty, breach, causation, damages. Each step must be supported by evidence grounded in the clinician's documented reasoning. But when the critical decision point was shaped—subtly or decisively—by an AI recommendation that cannot be interrogated, the causal chain collapses. The problem is compounded by the fact that many clinical AI tools are proprietary. Their code is protected as intellectual property. Their training data is secret. This creates a litigation asymmetry: the patient and physician may be harmed by an AI output, but neither can obtain the forensic tools needed to assess its validity.

Legal scholars and ethicists warn that such opacity may shield developers from meaningful scrutiny while exposing clinicians to disproportionate liability [26, 27]. The burden of proof, already steep in medical negligence claims, becomes insurmountable when the algorithmic pathway is not merely obscure but unavailable. This evidentiary breakdown also fractures the regulatory function of the tort system. Tort law is not only compensatory—it is corrective. It identifies negligence, assigns responsibility, and ideally deters future harm. But if harm caused by black-box systems cannot be reconstructed in legal terms, tort law can no

longer fulfill its regulatory role. The system becomes blind to the very errors it is meant to expose.

Some have proposed technical solutions such as explainable AI (XAI), which aims to generate human-readable rationales alongside predictions. While promising, these efforts often trade off performance for transparency. Moreover, regulators have yet to mandate explainability as a precondition for clinical deployment. Until they do, the use of black-box AI will continue to outpace the legal system's capacity to evaluate its effects. Another response is to shift the locus of liability toward enterprise actors—those who profit from the deployment of inscrutable tools—rather than holding clinicians accountable for systems they did not design.

In the absence of reform, the evidentiary logic of malpractice law may become increasingly symbolic: a courtroom drama in which the real decision-maker—the algorithm—is both silent and shielded. This is not merely a legal inconvenience. It is a structural threat to the legitimacy of the liability system itself.

The Erosion of Informed Consent and the Rise of Informed Refusal

The doctrine of informed consent rests on the presumption that a competent patient can receive, comprehend, and assess relevant information about their care options. But this legal pillar becomes unstable when the mechanisms driving care are not fully comprehensible to the physician, let alone the patient. Clinical AI systems, particularly those using opaque or evolving algorithms, challenge not only what it means to 'inform,' but also who is truly in a position to be informed.

Physicians have long served as interpreters of medical evidence, translating complex data into terms their patients can understand. But when the AI system becomes a de facto co-decision-maker—generating recommendations that are probabilistic, data-derived, and non-intuitive—this translation begins to fracture. Clinicians may not fully grasp how the AI reached its conclusion, particularly when black-box systems are involved. Even the most well-intentioned physician may therefore be unable to meet their duty to disclose risks or alternatives in a meaningful way [26,27].

This leads to a deeper rupture: if neither the physician nor the patient can interrogate the AI's logic, can informed consent truly occur? At best, consent risks becoming a symbolic gesture—an assent to a process whose inner workings remain hidden. At worst, it becomes a legal fiction, a ritual that obscures the absence of real human comprehensibility.

Equally important is the corollary principle of informed refusal. In traditional models of care, patients retain the right to reject a treatment they understand. Yet as AI becomes embedded into clinical workflows—whether through decision support alerts, diagnostic triage, or automation of routine tasks—this right erodes. Patients may never be told that AI shaped their care, or may not be offered a non-AI alternative. Refusal, in such contexts, becomes not only difficult but effectively impossible.

Scholars have proposed partial remedies. Some call for algorithmic traceability, where developers are required to provide post hoc explanations of decision paths. Others suggest tiered disclosure frameworks, where patients are informed not of the algorithm's technical logic but of its role, limitations, and level of autonomy [18]. Yet even these interventions may restore only the form of consent, not its substance. As Burrell (2016) argues, opacity is often not accidental but structural in machine learning, making true interpretability perpetually elusive [28]. Wachter, Mittelstadt, and Floridi similarly note that explainability may always lag behind performance, leaving transparency compromised in practice [17].

This legal uncertainty carries real consequences. As Sullivan and Schweikart contend, tort doctrines designed for human agency falter when applied to systems that resist explanation [11]. If neither clinician nor developer can map the causal chain of an AI-influenced decision, then the doctrines of foreseeability, causation, and fault risk collapsing into legal fictions.

The coming decade will require courts and policymakers to confront this gap directly. The principle of informed consent—like the standard of care—was forged in an era of human legibility. As that legibility fades, the law must decide: either reaffirm the centrality of human interpretability in medicine, or redefine consent in a world where interpretation itself has been outsourced to the machine.

The Role of the Hospital as Intermediary and Gatekeeper

In the evolving landscape of clinical AI, hospitals are no longer mere sites of care delivery. They are emerging as institutional intermediaries—entities that select, procure, deploy, and integrate AI systems directly into clinical workflows. This role positions the hospital not only as a conduit but also as a gatekeeper with both operational authority and potential legal exposure.

Historically, hospitals have been viewed as vicariously liable for the actions of their employees, including nurses and some physicians. But as more clinical decisions are shaped by algorithmic outputs, hospitals face direct liability for their role in system selection, training, and oversight. When an AI system is embedded into the electronic medical record, used as a triage or diagnostic aid, or incorporated into a robotic device, it is the hospital—not the individual clinician—that governs its presence and functionality.

Courts are beginning to recognize this institutional role. In Taylor v. Intuitive Surgical, Inc, the Washington Supreme Court held that manufacturers owe a duty to warn purchasing hospitals in addition to operating physicians, reasoning that hospitals must be informed of the risks of complex devices in order to credential physicians and provide adequate patient care [25]. Although that case did not involve AI, its logic is easily extended: hospitals that adopt high-risk AI tools without adequate safeguards or staff training may be seen as primary actors, not merely as venues.

The doctrine of the learned intermediary becomes strained in this environment. If the AI's functioning is opaque, if its integration is driven by institutional policy, or if the physician cannot meaningfully interrogate its logic, then the hospital becomes the last meaningful human intermediary in the chain of care. If that chain fails, so too does the rationale behind deflecting liability away from the institution.

Hospitals may also face non-delegable duties under certain theories of tort law. These include duties to ensure informed consent, provide adequate supervision, and maintain a safe clinical environment. When AI tools are deployed without full transparency, without appropriate audit mechanisms, or with biased outputs, these duties may be violated—even if no individual clinician acted negligently.

Moreover, the growing trend toward enterprise software bundles—where diagnostic AI, EMR tools, and scheduling algorithms are provided by a single vendor—further blurs institutional boundaries. The hospital may be purchasing not just tools, but an ecosystem of decision-makers, effectively outsourcing parts of its clinical governance to an opaque supply chain. As AI systems

become more embedded, the hospital increasingly occupies the hinge position—between machine and human, between policy and practice, between innovation and accountability.

Revisiting the Standard of Care in Algorithmic Medicine

The concept of the standard of care sits at the fulcrum of liability law. It is the benchmark against which negligence is measured, malpractice is judged, and professional conduct is defended. But as AI systems become embedded in clinical workflows, the very notion of what constitutes a 'reasonable' medical act is being rewritten—not by human consensus, but by machine optimization. Traditionally, the standard of care reflects the practices of a reasonably competent physician under similar circumstances, informed by clinical guidelines, peer practices, and evolving medical knowledge. It is backward-looking in form but normative in function: courts ask what a physician should have done, based on what others do. AI destabilizes this foundation. When an algorithm trained on millions of data points offers a recommendation that diverges from prevailing clinical norms, which standard applies—the human or the machine? If the physician follows the AI and a bad outcome result, is that negligence? If they ignore the AI and the patient is harmed, is that deviation?

Some commentators argue that a new 'algorithmic standard of care' is emerging. In this view, the use of AI becomes obligatory—not optional—once its superior performance is established in certain tasks. Failure to use AI could then constitute a breach [11]. This flips the liability equation: once AI becomes best practice, it becomes the legal baseline, and opting out is the riskier course. But this logic presumes stability, transparency, and broad access to AI tools—conditions rarely met in real-world clinical settings.

Equally troubling is the idea that AI could generate local or system-specific standards, contingent not on universal best practices but on the behavior of other algorithms in the same health system. If hospital A uses one AI and hospital B uses another, do their physicians owe different duties to their patients? This could fragment the standard of care into siloed, algorithmically derived practices that vary by vendor, training set, or geographic market.

As courts begin to confront these dilemmas, one thing becomes clear: the standard of care is no longer solely a human construct. It is being reshaped—subtly, unevenly—by systems that were not trained to think ethically, interpret context, or explain themselves. If left unexamined, this shift may result not in better care, but in a new legal opacity, where responsibility floats between code, clinician, and court, with no stable center to hold.

Proposed Solutions: From Enterprise Liability to Regulatory Innovation

As clinical AI erodes the boundaries between product and practice, and as traditional doctrines strain under the weight of algorithmic complexity, the legal system faces an inflection point. Simply retrofitting old liability models may no longer suffice. Instead, scholars and policymakers have proposed a range of structural reforms aimed at distributing accountability more fairly, incentivizing safety, and preserving innovation.

Enterprise Liability: Reassigning the Risk

One of the most widely discussed models is enterprise liability, which shifts responsibility away from individual clinicians and toward the organizations that develop, deploy, or profit from AI systems [26]. The logic is simple: hospitals and manufacturers are better positioned to assess system-level risks, implement safeguards, and absorb legal costs. Under this model, the physician remains a participant in care but is no longer the primary legal target when AI errors occur.

Regulatory Innovation: Learning from the EU and Beyond

The European Union's AI Act represents a landmark attempt to classify AI systems by risk level, impose transparency obligations, and define the duties of deployers, developers, and users [18]. High-risk medical AI tools are subject to conformity assessments, human oversight mandates, and post-market monitoring.

Though the U.S. lacks an equivalent comprehensive AI law, FDA initiatives like the Software Precertification Pilot Program and draft guidance on Good Machine Learning Practices point to a gradual shift. These frameworks aim to adapt existing medical device oversight to AI's continuous learning and software-based nature—but gaps remain in areas like post-deployment drift, adaptive behavior, and real-world harm attribution.

Insurance and Risk Pooling

Given the uncertainty of AI-related claims, insurance-based models are also gaining traction. Some have proposed AI-specific malpractice riders or enterprise-level insurance pools to absorb costs from unpredictable harms. This mirrors historical precedents, such as the National Vaccine Injury Compensation Program, which shifted litigation into a no-fault adjudicative model.

Reimagining Informed Consent and Informed Refusal

Legal reforms must also address the patient-clinician relationship. Some ethicists suggest reframing the duty not just as informed consent but as informed refusal—giving patients the right to opt out of AI-guided care when alternatives exist [27]. This acknowledges the epistemic asymmetry between human and machine reasoning while restoring a modicum of patient agency.

A Layered and Adaptive Legal Framework

No single solution will suffice. The future of clinical AI governance may require a layered approach—combining enterprise accountability, regulatory oversight, insurance mechanisms, and ethical practice reforms. Such a system would treat AI not as an external tool but as a coherent actor within a network of legal, technical, and clinical norms. Liability would become a shared, dynamic structure—not a blame game, but a safety net.

Historical Perspectives in Legal Adaptation

Artificial intelligence appears to confront us with new and bewildering challenges. Yet the problem of assigning responsibility to non-human agents, distributed causality, and technological opacity is not without precedent. Roman jurists and Islamic legal scholars wrestled with strikingly similar questions. Their solutions reveal that law has long adapted to novel forms of agency, often by inventing fictions, reinterpreting intention, and expanding the reach of responsibility.

Roman Law: Fiction and Foresight

Roman law provides some of the earliest and most enduring examples of how liability was extended beyond direct human action. The Digest of Justinian declared that the keeper of an animal bears liability for damage it causes, even in the absence of personal fault [29]. Here, responsibility rested not on the internal intent of the animal, but on the human who introduced the risk into the world. Agency did not require consciousness; it required causation.

The Lex Aquilia, drafted in the third century BCE to address property damage, soon became a vessel for broader interpretations. The Roman jurist Ulpian extended its reach to consider not only what was done, but what ought to have been foreseen—a significant step toward anticipatory liability [30]. As legal historian Peter Stein later reflected, Roman jurists employed fictio juris—legal fictions—not as deceptions but as bridges, allowing the law to reach situations it could not yet name [31]. Such fictions resonate

with today's struggles over AI: conceptual devices to link opaque, distributed systems back to accountable actors.

Islamic Jurisprudence: Intention, Tools, and Delegated Agency
Islamic jurisprudence developed parallel mechanisms, particularly in addressing indirect causation (tasabbub). Scholars carefully distinguished between immediate acts and mediated harm, often locating responsibility in the initiation of the causal chain rather than its conclusion. In the classical tradition represented by scholars such as Al-Sarakhsi, responsibility did not evaporate simply because harm was produced through an instrument rather than by a direct hand—a principle with direct bearing on algorithmic liability today [32].

Islamic legal reasoning recognized that those who set a process in motion bear responsibility for its consequences, even when the mechanism operates long after the initiating act. Ibn Rushd, drawing on both Aristotelian philosophy and Islamic jurisprudence, argued that legal accountability must extend into domains that exceed ordinary comprehension—a principle that speaks directly to the challenge of governing black-box AI [33].

Echoes Across Traditions
Roman and Islamic jurisprudence converge on a central lesson: legal systems do not require human sentience to impose responsibility. They recognized that risk originates in deployment, not merely in execution. Whether through the Roman bridge of fictio juris or the Islamic doctrine of delegated causation, both traditions crafted frameworks to address agency when human control seemed distant. What remains constant is the demand that accountability never vanish into opacity. That jurisprudential wisdom is precisely what must guide us in the age of artificial intelligence.

Conclusion: Carrying Forward the Jurisprudential Echoes
The lessons of Roman and Islamic jurisprudence remind us that the law has never been static. It has always been asked to stretch across gaps in human control, to impose accountability where causation is indirect, and to guard against opacity. As Peter Stein observed of the Roman legal tradition, the fiction is not a lie, but a bridge [31]. That bridge now leads us into the era of artificial intelligence.

Today, we face not animals or irrigation wheels but neural networks and autonomous clinical tools. The parallels are striking: law must again confront agents that act beyond the hand of their creators. The Romans taught us to anticipate harm through foresight; Islamic jurists taught us that the one who sets a process in motion must bear its consequences. Together, they warn us that responsibility must not vanish simply because the mechanism is complex.

Yet history also reveals the cost of delay. The cycles of innovation in medicine show that caution is too often eclipsed by the pressure to deploy first. From Theranos to IBM's Watson for Oncology, unfinished tools have been unleashed before safeguards were ready, leaving patients to absorb the harm. Even Sam Altman, CEO of OpenAI, acknowledged before the U.S. Senate in May 2023: 'I think if this technology goes wrong, it can go quite wrong, and we want to be vocal about that' [15].

The path forward requires both humility and anticipation. Tort law, with its adaptability, can absorb AI into its fabric—through doctrines of strict liability, enterprise liability, or an evolved learned intermediary role. But these adjustments must happen now, not retroactively after preventable harm has occurred.

Artificial intelligence in healthcare is not ungovernable. It is the latest chapter in a story that stretches back to Rome and Baghdad, to the jurists who recognized that tools can outpace comprehension, yet still demand accountability. The challenge for our generation is to ensure that the law runs quickly enough this time—so that patients are not left as casualties in the space between invention and responsibility.

References

1. Pasquale F (2015) The black box society: The secret algorithms that control money and information. Harvard University Press.
2. Herper M (2018) The rise and fall of Theranos: A timeline. Forbes. https://www.forbes.com/sites/matthewherper/2018/03/14/the-rise-and-fall-of-theranos-a-timeline/.
3. Ross C, Swetlitz I (2017) IBM pitched its Watson supercomputer as a revolution in cancer care. It's nowhere close. STAT News. https://www.statnews.com/2017/09/05/watson-ibm-cancer/.
4. Carreyrou J (2018) Bad blood: Secrets and lies in a Silicon Valley startup. Knopf.
5. Strickland E (2019) IBM Watson, heal thyself: How IBM overpromised and underdelivered on AI health care. IEEE Spectrum https://spectrum.ieee.org/how-ibm-watson-overpromised-and-underdelivered-on-ai-health-care.
6. U.S. Securities and Exchange Commission (2018) Theranos, CEO Holmes, and former president Balwani charged with massive fraud. https://www.sec.gov/news/press-release/2018-41.
7. Carreyrou J (2015) Hot startup Theranos has struggled with its blood-test technology. The Wall Street Journal. https://www.wsj.com/articles/theranos-has-struggled-with-blood-tests-1444881901.
8. Ross C, Swetlitz I (2018) IBM's Watson supercomputer recommended 'unsafe and incorrect' cancer treatments. STAT News. https://www.statnews.com/2018/07/25/ibm-watson-recommended-unsafe-incorrect-treatments/.
9. Cavallo J (2019) Watson disappoints. Cancer Today. https://www.cancertodaymag.org/spring2019/watson-disappoints/.
10. RAND Corporation (2024) Liability for harms from AI systems: The application of U.S. tort law and liability to harms from artificial intelligence systems. https://www.rand.org/pubs/research_reports/RRA3243-4.html.
11. Sullivan HR, Schweikart SJ (2019) Are current tort liability doctrines adequate for addressing injury caused by AI? AMA Journal of Ethics, 21(2), E160–166. https://doi.org/10.1001/amajethics.2019.160.
12. Griffin J (2021) Artificial intelligence and liability in health care. Health Matrix: The Journal of Law-Medicine, 31(1), 65. https://scholarlycommons.law.case.edu/healthmatrix/vol31/iss1/5/.
13. Duffourc MN, Gerke S (2023) Generative AI in health care and liability risks for physicians and safety concerns for patients. JAMA, 330(4), 313–314. https://doi.org/10.1001/jama.2023.9630.
14. Dolfing H (2024) Case study: IBM Watson for Oncology failure. https://www.henricodolfing.com/2024/12/case-study-ibm-watson-for-oncology-failure.html.
15. Altman S (2023). Written testimony before the U.S. Senate Judiciary Subcommittee on Privacy, Technology, and the Law. https://www.judiciary.senate.gov/committee-activity/hearings/oversight-of-ai-rules-for-artificial-intelligence.
16. Wallach W (2015) A dangerous master: How to keep technology from slipping beyond our control. Basic Books.

17. Wachter S, Mittelstadt B, Floridi L (2017) Transparent, explainable, and accountable AI for robotics. Science Robotics, 2(6), eaan6080. https://doi.org/10.1126/scirobotics.aan6080.
18. European Parliament & Council of the European Union (2024) Artificial Intelligence Act (AI Act), Regulation (EU) 2024/1689. https://artificialintelligenceact.eu.
19. Cohen IG, Spector-Bagdady K (2022) The promise and perils of evidence generation in learning health systems. Journal of Law and the Biosciences 9: lsac001.
20. Price WN, Gerke S, Cohen IG (2019) Potential liability for physicians using artificial intelligence. JAMA 322: 1765-1766.
21. American Law Institute (1965) Restatement (Second) of Torts. American Law Institute Publishers.
22. Keeton WP, Dobbs DB, Keeton RE, Owen DG (1984) Prosser and Keeton on the law of torts (5th ed.). West Publishing.
23. American Law Institute (1998) Restatement (Third) of Torts: Products Liability. American Law Institute Publishers.
24. Wagner WE (2007) When all else fails: Regulating risky products through tort litigation. Georgetown Law Journal 95: 693.
25. Taylor v (2017) Intuitive Surgical, Inc., 389 P.3d 517, 187 Wash. 2d 743 (2017).
26. Price WN, Gerke S, Cohen IG (2021) Potential liability for physicians using artificial intelligence. JAMA 325: 177-178.
27. Gerke S, Minssen T, Cohen IG (2020) Ethical and legal challenges of artificial intelligence–driven healthcare. The Hastings Center Report 50: 40-43.
28. Burrell J (2016) How the machine 'thinks': Understanding opacity in machine learning algorithms. Big Data & Society 3: 1-12.
29. Watson A (1998) The Digest of Justinian (2 vols., rev. ed.). University of Pennsylvania Press.
30. Buckland WW (1932) A textbook of Roman law from Augustus to Justinian (2nd ed.). Cambridge University Press.
31. Stein P (1999) Roman law in European history. Cambridge University Press.
32. Hallaq WB (2009) An introduction to Islamic law. Cambridge University Press.
33. Gutas D (2001) Avicenna and the Aristotelian tradition: Introduction to reading Avicenna's philosophical works. Brill.