Crimson Publishers
Wings to the Research

Review Article

From Transparency to Torts: Navigating Legal Liability in Healthcare AI

Andrew J Fishman MD[1,2]*

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

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

ISSN: 2832-4463

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

Submission: August 31, 2025
Published: October 13, 2025

Volume 5- Issue 1

How to cite this article: Andrew J Fishman MD*. From Transparency to Torts: Navigating Legal Liability in Healthcare AI. COJ Rob Artificial Intel. 5(1). COJRA. 000602. 2025.
DOI: 10.31031/COJRA.2025.05.000602

Abstract

Artificial Intelligence (AI) is transforming the landscape of clinical care, offering tools that range from decision support to fully autonomous diagnosis and treatment. Yet this innovation has outpaced the development of clear legal frameworks. Traditional tort law and product liability doctrines strain to accommodate a world in which responsibility for harm may be buried within opaque algorithms or distributed across an intricate web of systems, developers, and institutions. This review examines the emerging legal landscape applicable to healthcare AI solutions, comparing U.S. and EU regulatory approaches and drawing on key case law to explore liability boundaries. We argue that tort law is adaptable to the AI revolution in healthcare. But this evolving legal framework only works if courts, developers, healthcare enterprises, and clinicians acknowledge the shifting terrain of transparency, autonomy, and accountability.

Keywords: Artificial intelligence; European union; Food & drug administration: Healthcare; Law; Legal; Medicine; Robotics; Surgery

Introduction: Law Meets the Algorithm

Artificial Intelligence (AI) has rapidly become an integral component of modern healthcare, assisting clinicians around the world with data-driven tasks. Recent surveys reflect this surge: about two-thirds of U.S. physicians now report using AI-based tools for documentation, discharge planning, translation, and diagnostic support [1]. Additionally, nearly 80% of U.S. healthcare organizations report that they are implementing AI technologies in their operations [2]. Globally, similar momentum is evident: one international survey found that 48% of clinicians have used AI tools in practice (nearly double the share from the previous year) [3]. Despite these strides, the World Economic Forum underscores that healthcare still trails other industries in AI maturity and deployment [4].

These AI-powered systems span the spectrum from predictive risk models for patient triage or sepsis alerts to advanced imaging interpretation and even AI-enhanced robotic surgical assistants and ambient listening documentation assistants. They promise significant gains in diagnostic accuracy, efficiency, and access to care. Yet as these systems increasingly influence clinical outcomes, their deployment also raises a critical legal question: if an AI-informed clinical decision leads to patient harm, who bears responsibility? Is it the physician who relied on the output? Did the hospital successfully integrate the tool into the clinical workflow? Or should the developers who trained the AI on incomplete data be held responsible? Some legal scholars have even proposed liability be assigned to the AI itself by imparting personhood status.

Most physicians, developers, and engineers are trained to think in terms of accuracy, data integrity, and systems performance, not liability. But as AI increasingly mediates high-stakes clinical judgments, the legal system must answer a difficult question: can existing frameworks for medical negligence and product liability accommodate software that "thinks" but cannot explain how it does so?

Consider a practical case. A 47-year-old man presents to the emergency department with chest pain. The triage AI flags him as low priority based on its analysis of presenting symptoms and vital signs. Hours later, he suffers a myocardial infarction. Retrospective analysis reveals that the AI had never been trained on the specific comorbidity pattern this patient presented. Who failed him? Was it the physician who trusted the tool, the hospital that installed it, or the developer that designed the training set?

This article offers a roadmap for clinicians and technologists seeking to understand the legal terrain that surrounds the use of AI in medicine. It assumes no formal legal training and is premised on the following set of core definitions derived from U.S. common law principles:

A. Tort law is the branch of civil law that governs harm caused by one party to another. Medical malpractice and product liability both fall under this domain [5].

B. Negligence is a fault-based tort claim that asks whether a person or institution failed to meet a reasonable standard of care [6]. Medical malpractice and general negligence are adjudicated according to this principle.

C. "Standard of care" refers to the level and type of care that a reasonably competent professional, in a similar role and under similar circumstances, would provide to a patient. In medical malpractice cases, this is predominantly established through expert testimony, and it forms the benchmark against which negligence is assessed [7].

D. Strict product liability applies to defective products regardless of fault, holding the manufacturer liable simply because the product was unsafe [8].

E. Causation is a required element in both negligence and product liability claims: it must be shown that the defendant's conduct (or product) directly caused the harm [6,9].

In tort law, a negligence claim requires proof of four essential elements: duty, breach, causation, and damages. First, the defendant must have owed a duty of care to the plaintiff under the circumstances. Second, the defendant must have breached that duty by failing to act as a reasonably prudent person would have acted. Third, there must be a direct causal link between the breach of a duty of care and the harm suffered. Finally, the plaintiff must have sustained actual damages, whether physical, financial, or otherwise, because of the defendant's actions or omissions. To prevail on a negligence claim, the plaintiff must also establish causation, both factual and legal. Factual causation, often called "but-for" causation, asks whether the harm would have occurred but for the defendant's conduct. Legal causation, or proximate cause, limits liability to those harms that are reasonably foreseeable consequences of the act or omission. Together, these elements ensure that liability is imposed only when the defendant's breach was both a necessary condition and a legally cognizable cause of the injury.

In strict product liability, causation and injury remain essential elements, but the focus shifts away from duty and breach toward the presence of a product defect. The plaintiff must show that the product was defective and that this defect directly caused the harm. A product may be defective in its design, meaning it is inherently unsafe as intended, even if manufactured correctly. Additionally, a product may be considered defective if there is a manufacturing flaw, which occurs when an error in the production process causes the product to differ from its intended design. Finally, a product can be defective if it lacks adequate warnings or instructions about known or foreseeable risks. In this framework, liability is based on the condition of the product itself, not on whether the manufacturer acted negligently.

Let's go back to the previous example where an AI tool incorrectly triaged the patient. While that scenario is illustrative, it reflects real and growing concerns in the healthcare industry today. The experience with Electronic Health Records (EHRs) provides a historical parallel. EHRs were widely expected to reduce errors and streamline care, yet in practice they often increased clinician workload, introduced new failure modes, and became sources of precedent-bearing litigation, which we will review in this article. A 2023 JAMA review underscored how EHRs contributed to physician burnout, alarm fatigue, and workflow fragmentation, despite their foundational role in digital healthcare [10].

AI increasingly repeats patterns with deeper complexity. As a 2025 BMC Medical Ethics study observes, clinicians worry that AI's "black box" nature, especially in deep neural networks and complex models, often renders systems unexamined and their internal logic inaccessible, thereby distancing clinicians from the reasoning that underpins clinical decisions [11]. This creates a dangerous ambiguity: if a physician relies on an AI recommendation that cannot be interrogated, how should a court weigh accountability?

Moreover, many AI systems now exhibit both autonomy and adaptiveness, evolving beyond their initial programming and making or influencing decisions with limited human oversight. Tools such as robotic surgical platforms, autonomous triage software, and self-updating diagnostic models act within clinical workflows as "functionally independent actors" that can influence outcomes without possessing legal personhood, intent, or duty. As noted by the European Parliamentary Research Service, this blurring of agency challenges existing legal frameworks of accountability and redress [12]. The central concern is not simply who to blame, but how to ensure patient safety and maintain trust in a healthcare system undergoing rapid technological transformation.

The U.S. and EU are beginning to diverge in their approaches to this issue. While the U.S. still relies heavily on judicial doctrines

and FDA approval pathways, the EU has moved forward with comprehensive legislative reforms, including the AI Act and updates to liability law through its revised Product Liability Directive (PLD). These efforts redefine software as a product and introduce structural presumptions that reshape liability dynamics. Under the proposed AI Liability Directive (which was eventually withdrawn in advance of adoption of the AI Act), the EU considered establishing a rebuttable presumption of causality, easing the burden of proof for plaintiffs when AI systems cause harm [13]. Furthermore, the new PLD (Directive EU 2024/2853) explicitly includes software, including standalone AI, as a "product," expands potential defendants across the supply chain, and enables courts to presume both defectiveness and causation in complex cases of digital product malfunction [14,15]. These combined reforms significantly shift the legal landscape, tilting the balance toward consumer protection when AI-inflicted harm occurs.

This article is intended to serve as both a legal primer and a practical guide. It unpacks the evolving legal standards governing AI in medicine, examines how liability is being redistributed across physicians, institutions, and developers, and outlines pathways proposed by scholars and policymakers to promote fairer and more effective accountability. While this redistribution remains more pronounced in the EU than in the U.S., it reflects a global legal landscape in flux. Each section of this article builds a foundation for the next, beginning with core concepts such as explainability and autonomy and culminating in systemic proposals that bridge the gap between innovation and justice.

Methods and Scope

This article is a narrative review that synthesizes legal, regulatory, and scholarly sources relevant to artificial intelligence in healthcare. The analysis draws primarily on U.S. and European Union frameworks, including tort law, product liability doctrines, and regulatory guidance from the U.S. Food and Drug Administration (FDA), the EU Artificial Intelligence Act (Regulation [EU] 2024/1689), and the revised EU Product Liability Directive (Directive [EU] 2024/2853). Case law examples from U.S. courts and policy documents from the European Parliament and Commission are incorporated to illustrate liability boundaries and emerging trends. Peer-reviewed literature from medical, legal, and ethics journals provides additional context. The scope is limited to the comparative examination of U.S. and EU jurisdictions, with international references included where they clarify general principles or future directions [16].

Opacity, Evidence, and Accountability

A foundational principle of modern tort law is that harm must be traceable. Whether one is alleging negligence (a failure to act reasonably) or a product defect (a flaw in design or manufacture or a failure to warn), the legal system depends on reconstructing what happened and why to establish causation. A core technical reality in the realm of artificial intelligence, particularly machine learning, immediately challenges this principle because many AI systems lack interpretability. Even when their outputs appear reliable, the inability to scrutinize underlying logic in cases of suspected error creates a serious evidentiary barrier to accountability.

Any discussion of AI liability must begin with transparency, or more precisely, the absence of it. In traditional medical malpractice litigation, the chain of reasoning can be reconstructed. One can examine the physician's notes, consult clinical guidelines, and compare decisions made by the treating physician to established standards. But when AI enters the equation, especially black box systems, this forensic trail can vanish. While interpretability tools, such as post hoc explanation algorithms, exist to approximate a model's internal logic, they do not reveal the actual reasoning process. At best, they provide simplified estimates of influential factors, which may not hold up under legal scrutiny.

This opacity, often referred to as the black box problem, refers to AI models, especially deep learning systems, that produce results without providing a human-readable explanation for how those results were derived. In contrast, white box models are fully transparent. They follow fixed, rule-based structures with predictable and traceable logic. Gray box models sit in between, offering limited insight into contributing variables but not a complete map of causation.

Several practical examples illustrate the gradient of transparency:

a) **White box AI:** A clinical decision support tool flags a potential drug-drug interaction using a fixed ruleset. The clinician can review the flagged medications and the rule that triggered the alert.

b) **Gray box AI:** A sepsis risk prediction model assigns a high-risk score based on a combination of vitals, labs, and clinical notes, with partial visibility (e.g., top contributing features).

c) **Black box AI**: A convolutional neural network scans a CT and outputs "no abnormality" without identifying what it looked at or how it reached that conclusion.

Transparent AI models, both white box and gray box, have long supported clinical decision-making with traceable logic. By contrast, truly opaque "black box" systems, such as deep learning models in medical imaging, are a more recent development and remain uncommon in practice. In the white box case, courts can evaluate whether the model was properly configured or whether the clinician ignored an obvious alert. In the black box case, neither the court nor the clinician may ever know what the system "saw" or ignored.

To date, no U.S. court has issued a clear ruling against a black box AI; in contrast, white- and gray-box models have already featured in precedent-setting malpractice cases that hinge upon traceable algorithmic reasoning. However, when a black box tool produces results such as "no abnormality" without providing any explanation, clinicians and courts are left without any insight into what the system "saw" or omitted, compounding the legal burden of proving causation under tort law [17].

The legal implications of this spectrum are profound. When a physician makes a mistake, courts can evaluate their reasoning against the standard of care. However, the decision chain becomes obscure when a black box AI system recommends a diagnosis, or fails to recommend one, and the physician acts on it. The AI system may have factored in countless variables in a manner no human can replicate or audit. As such, causation becomes speculative, undermining one of the pillars of tort liability. Recent commentary from the 2025 Stanford Technology Law Review articulates this issue with precision: "Black box systems challenge the evidentiary logic of liability law. They create clinical consequences without evidentiary trails." This has practical ramifications. If an AI diagnostic system misses early signs of a stroke or recommends a contraindicated drug interaction, plaintiffs may struggle to prove that the harm was caused by a flaw in the algorithm, or that any human should have known better [18].

The issue is compounded by the fact that many AI systems are trained on proprietary datasets and use closed-source architectures. This limits not only clinical interpretability but also judicial discovery. Courts cannot compel transparency if the model's internal workings are not disclosed, nor can expert witnesses opine effectively on systems they cannot review. The European Parliament's 2020 report on AI in healthcare warned of this explicitly: "Without mandated transparency, AI introduces accountability gaps that cannot be resolved post hoc" [12]. The legal system has not yet resolved this dilemma. As of 2024, U.S. case law does not mandate explainability for AI systems used in clinical settings. Nor do FDA approval pathways such as 510(k) or De Novo currently mandate explainability as a condition of clearance. As a result, hospitals and physicians may adopt highly accurate but inscrutable AI tools without knowing they are also taking on unquantifiable legal risk. Furthermore, there are signals in the US court system that attributing fault solely to an AI system is unlikely to be a viable defense. In Skounakis v. Sotillo [19] (N.J. Super. Ct. App. Div. 2018), a New Jersey appellate court addressed in an unpublished decision the use of clinical software in guiding physician decisions after a patient died following a prescribed combination of phendimetrazine and liothyronine (Cytomel) for weight loss. Although the software was not a black box system, the case is instructive in delineating the limits of liability when algorithmic recommendations are involved. The trial court had initially excluded the plaintiff's expert, who was a cardiologist rather than an OB/GYN and not a software engineer. This effectively left the plaintiff without a qualified causation witness. However, the appellate court overturned the lower court's summary judgment ruling in favor of the defendant, concluding that the physician's alleged breach of the standard of care had received sufficient expert testimony from the plaintiff's expert to consider the negligence claim on the merits. This allowed the case to proceed to trial, underscoring that physicians remain potentially accountable even when relying on software tools [19,20].

A 2025 BMC Medical Ethics report warns that lack of interpretability of AI tools also undermines informed consent. If a clinician cannot understand how the AI system reaches its conclusions, how can they responsibly discuss its risks and limitations with the patient? [21] From a policy perspective, the EU AI Act offers a more stringent approach. For high-risk healthcare AI systems, the Act mandates not only technical documentation and risk management protocols but also explainability "proportionate to context." While not demanding full transparency for black box systems, the EU's position is that opacity must be counterbalanced by robust human oversight and disclosure standards [16]. The divergence in regulatory philosophy between the U.S. and EU is becoming clear. The U.S. has placed emphasis on market deployment, while the EU prioritizes precaution and oversight.

The shifting standard between explainability and accountability will shape both litigation and design practices in the years to come. Developers, regulators, and clinicians must understand that transparency is not just a design choice; it is a legal defense. And in its absence, new legal frameworks for allocating responsibility must be built.

Adaptivity: A New Axis of Legal Complexity

Another crucial and often overlooked dimension of medical AI is whether a system is fixed (or "locked") or adaptive. This distinction has far-reaching consequences not only for clinical behavior but also for legal responsibility. A fixed AI system is trained on a defined dataset and remains unchanged after deployment. Its decision-making parameters are static, and any updates require a new review cycle or formal revalidation. This model resembles traditional software tools or medical devices. In contrast, an adaptive AI system continues to evolve, ingesting new data and adjusting its parameters dynamically. These systems are "self-learning" in the sense that their outputs may change over time without explicit reprogramming.

This adaptive behavior offers substantial clinical promise, real-time responsiveness to emerging data, personalization across populations, and iterative performance improvement. However, the gradual shift in a model's outputs over time may occur in ways that are difficult to anticipate, audit, or retrace. From a legal standpoint, this poses a fundamental challenge. In the case of fixed systems, courts can evaluate the system as it existed at the time of harm. But with adaptive systems, the model that produced the output may no longer exist in the same form by the time of litigation. Because the algorithm has shifted, the same patient presenting with identical symptoms might receive different recommendations in January versus March. Unless full logging of changes, historical data and algorithm performance capture, and delineable version control are maintained, causation becomes a moving target. Adaptive AI further complicates this by introducing continuous, rather than discrete, performance changes-making it difficult to pinpoint when a new risk became known or actionable. This creates heightened exposure under failure-to-warn theories in product liability, where evolving evidence (such as newly published adverse findings) may emerge gradually, challenging the timeliness and adequacy of disclosures.

The lack of system permanence destabilizes traditional notions of liability and presents a dual evidentiary challenge: plaintiffs may struggle to establish causation because the exact version of the AI that generated the harmful recommendation may no longer exist. Meanwhile, defendants may find it equally difficult to mount a defense, particularly if the system's performance drifted due to inadequate post-deployment monitoring Chew et al. [22]. This erosion of traceability strikes at the core assumption of product liability: that the product causing the harm can be examined, tested, and judged. Adaptive systems complicate this. Consider the practical implications: a patient develops severe sepsis after an AI system failed to generate an alert. During discovery, the model is found to have been retrained three times since the event. What evidence is relevant and admissible? Which version is culpable?

To address these challenges, the U.S. Food and Drug Administration (FDA) has proposed a Total Product Lifecycle (TPLC) framework for approvals of AI/ML-based Software as a Medical Device (SaMD) product. This approach treats AI systems as "living products," requiring Predetermined Change Control Plans (PCCPs), ongoing performance and safety monitoring, and transparent documentation of all training and retraining events [23]. The goal is to ensure that updates do not compromise safety or efficacy. Yet, as of 2024, implementation of products under this framework remains limited, and no federal mandate requires healthcare institutions to audit adaptive behavior post-deployment.

Most hospitals deploying adaptive AI tools do not maintain structured systems to monitor how these tools evolve over time, resulting in accountability gaps that often only surface after patient harm has occurred. As recently reported by KFF Health News, "many institutions are not routinely monitoring the performance" of these AI products-and, according to then FDA Commissioner Dr. Robert Califf, "I do not believe there's a single health system in the United States that's capable of validating an AI algorithm that's put into place in a clinical care system" [24].

In the EU, the approach is far more definitive. The 2024 EU AI Act mandates ongoing risk management systems for adaptive models, emphasizing transparency, robustness, and sustained accuracy. Crucially, "AI systems intended to be used as safety components in the management and operation of critical digital infrastructure and life-critical environments, including healthcare, shall be considered high risk" [25]. As a result, healthcare AI systems are required to undergo third-party conformity assessments and must maintain detailed audit trails documenting model updates-making the regulation both clear and enforceable [16].

These regulatory frameworks implicitly acknowledge that adaptive systems blur the line between product liability and ongoing professional or institutional duty. If a system evolves after deployment, does liability rest with the original developer, the deploying physician or institution, or the team responsible for retraining and integration? This was illustrated in the 2023 Science review on AI in translational medicine, which warned, "Without a clearly defined boundary between system designer, deployer, and operator, the chain of accountability dissolves under pressure" [26].

To manage risk, institutions deploying adaptive AI systems should establish local documentation procedures that clearly record the sources of training data and any model update events. Internal ownership must be assigned for model validation and ongoing audit responsibilities, ensuring that accountability is not diffused across departments or vendors. In addition, systems should include patient safety flags or rollback capabilities to respond to known deviations or performance drifts.

In short, adaptivity is not merely a technical feature, it is a legal hazard. It creates a moving target for liability, one that neither tort law nor existing FDA or EU regulation has yet fully addressed. Until clearer legal standards emerge, adaptivity remains a key source of uncertainty, demanding proactive oversight from clinicians, developers, and institutions alike.

Autonomy and the Fracturing of the Standard of Care

In tort law, particularly in medical malpractice, the standard of care refers to the level of competence and diligence that a reasonably skilled healthcare provider is expected to exercise under the same or similar circumstances. For decades, this standard has been defined by professional norms, specialty guidelines, and evolving clinical knowledge-and has been established in court primarily through expert testimony. Physicians are expected to evaluate available evidence, apply their training, and exercise judgment, an inherently human-centered framework. This framework presumes human agency. It assumes that a person, not a machine, has evaluated the evidence, made a decision, and can explain it afterward. But artificial intelligence challenges every part of that presumption.

The introduction of autonomous AI systems in healthcare complicates this picture. When decisions are influenced or entirely made by AI, the line blurs between human judgment and machine reasoning. A critical legal question emerges: can a clinician be held liable for a decision made by an AI system they did not fully control or understand? To answer this, it is essential to distinguish between two categories of AI integration. Decision-support AI refers to systems that provide recommendations or augment physician reasoning but require human sign-off-such as risk scores or diagnostic prioritization tools. In contrast, autonomous AI includes systems that make or execute decisions independently, such as pathology department algorithms that automatically process specimens, finalize reports for normal cases, and forward only abnormal findings for human review.

In the first scenario, liability generally follows traditional malpractice logic. The physician remains the final authority and is judged against their duty to review, question, and interpret AI-generated outputs. Though there have yet to be precedential decisions issued specifically with respect to the responsibility of an AI system, there is already some legal guidance that addresses this issue. In Skounakis v. Sotillo [19], an OB/GYN relied on software-generated recommendations for weight-loss medication. The patient died, and although the software was implicated, the court reinstated negligence claims against the physician for failing to

exercise independent clinical judgment [19]. The tool augmented decision-making, but it didn't replace physician responsibility. Importantly, this software was not opaque in the sense of deep learning or AI-driven black-box models. Instead, its internal logic was presumably accessible or traceable, but the court emphasized that software recommendations do not replace the physician's duty to exercise independent clinical judgment.

With an autonomous AI system, however, legal responsibility shifts. In principle, the physician may play no direct role in the contested decision. Consider an AI-driven triage system that can direct patients to clinical care outcomes before any clinician is involved. Alternatively, robotic surgery platforms that remain under the global control of a surgeon may incorporate AI-modulated features to refine or stabilize instrument movements in real time. While the surgeon directs the overall procedure, certain micro-adjustments, such as tremor reduction, motion scaling, or trajectory smoothing-are autonomously tuned by embedded AI algorithms to enhance precision. Here too, there has been some legal precedent, again not directly focused on the AI's responsibility but rather on the robotics system itself. In Taylor v. Intuitive Surgical, Inc. (389 P.3d 517, Wash. 2017), the Washington Supreme Court addressed the liability of the manufacturer of the da Vinci® Surgical System to warn the hospital that purchased the device about associated risks. The court ruled that the device manufacturer owed a duty not just to the surgeon but also to the purchasing hospital, broadening the notion of who may be responsible for ensuring safe integration of complex technologies [27]. This reflects a growing legal trend: courts recognize that with autonomous tools, responsibility must be institutional, not just individual.

With autonomous AI, the traditional legal doctrine of the learned intermediary begins to unravel. Historically, this doctrine held that manufacturers could discharge their duty to warn by informing the physician, who would in turn counsel the patient about risks. But with AI-driven tools making independent decisions, the physician may no longer be the logical intermediary. The AI system itself becomes a de facto actor in the care chain, but it lacks legal personhood, intent, or accountability. The 2025 DePaul Law Review explores this erosion in detail: "The physician is no longer the only mind in the room. In black box AI, clinical responsibility becomes shared but legally orphaned" [28].

Shared causality without shared liability is one of the greatest structural risks in AI-enabled medicine. This is especially true in high-volume or fast-paced clinical care settings. Autonomous triage systems, telemedicine chatbots, and back-end prioritization tools are intended to manage patient flow often without direct physician oversight. U.S. courts have yet to fully confront the legal challenges posed by clinical AI. In analogous cases involving other medical technologies, product liability claims are often dismissed on preemption grounds-especially when the device has FDA Premarket Approval (PMA), a rigorous approval pathway that generally shields manufacturers from state law product liability claims. Should future AI tools follow this route, similar defenses may apply. However, most AI systems enter the market through less burdensome pathways such as 510(k) or De Novo classification, which do not offer the same legal protections. As a result, these faster approvals may expose developers to novel and still-evolving liability risks.

Meanwhile, the EU's AI Act attempts to close this gap by treating autonomous AI as high-risk, requiring traceability, post-market surveillance, and explicit human oversight requirements for deployment. (European Commission, 2024) Ultimately, the legal system must recognize that AI autonomy is not just a software feature-it is a jurisdictional fault line. When human oversight and control dissolves or is diminished, so too does the clarity of accountability. If clinicians are to remain legally responsible, they must retain a meaningful veto over AI recommended actions. Otherwise, new frameworks assigning liability to institutions or developers may be imposed to protect patients and ensure fair adjudication.

Software Becomes a Product: Tort Law in Transition

In U.S. tort law, one of the most consequential legal distinctions is between negligence-based liability and strict product liability. Negligence focuses on the conduct of individuals or institutions. Did they act with reasonable care? Strict product liability, by contrast, concerns the condition of the product itself. Was there a defect? For decades, software was treated more like a service than a product. This legal framing shielded software developers and vendors from the full force of product liability law. But with the rise of AI-based algorithms being deployed in clinical settings courts are beginning to reconsider this categorization.

That shift came into sharp focus in Lowe v. Cerner Health Services [29]. In that case, a patient suffered catastrophic brain injury after postoperative oxygen monitoring was delayed. The Electronic Health Record (EHR) system had defaulted the pulse oximetry start time to 10:00 a.m. the following day, rather than immediately initiating continuous monitoring when the doctor input the order as intended. The plaintiff alleged that this software configuration directly contributed to the harm. The court allowed the claim to proceed under product liability theories, marking a critical departure from the view that EHR software is a mere informational service [29]. By treating the EHR as a product rather than a service, the court opened the door to strict liability claims, meaning plaintiffs no longer had to prove negligence, only that the software was defectively designed or failed to carry adequate warnings. Restatement (Third) of Torts: Products Liability [6] Note that though widely quoted, this 4th Circuit decision is designated as unpublished, signaling that the opinion does not have precedential value and may be subject to citation restrictions depending on jurisdiction. But the direction is clear.

This potential legal reclassification of software enables courts to apply the three core theories of product liability to AI systems. First, design defect claims may arise when the AI system algorithm is inherently flawed or if it was trained on inappropriate or unrepresentative data. Second, manufacturing defect theories could apply if the deployed version of the AI system

was corrupted, improperly implemented, or poorly integrated into the clinical environment. Third, failure-to-warn claims may be triggered when parties in the chain of distribution, such as developers, manufacturers, or distributors fail to adequately disclose known limitations, biases, or risks associated with the system's performance. Failure to warn claims become even more problematic when deployed systems are adaptive as discussed earlier.

This aligns with evolving EU doctrines under the revised Product Liability Directive (PLD) and AI Liability Directive (AILD), both of which explicitly include standalone software within the definition of a "product." More significantly, the EU has introduced presumptions of causation for harm caused by high-risk AI systems, including medical AI. Under this directive, if a developer fails to produce adequate documentation of the system's design and development, courts may presume causation and shift the burden of proof to the developer or other responsible entity, rather than the plaintiff [30]. This marks a fundamental shift in legal exposure. Under strict liability, procedural diligence alone is not a defense. A developer may follow all industry standards, conduct thorough testing, and document every step yet still be held liable if the product is found to be defective and causes harm. The focus is not on whether the developer acted reasonably, but whether the product functioned safely. This standard lowers the evidentiary burden for plaintiffs and broadens the scope of litigation, particularly for adaptive systems whose behavior may change over time. In such cases, post-deployment evolution can introduce latent defects, exposing developers to liability even without negligence.

To counterbalance the recent plaintiff-friendly shift in strict liability adjudication, there is a parallel trend in the United States that moves in the opposite direction. In most U.S. jurisdictions, plaintiffs bringing strict liability design-defect claims must not only demonstrate that a product was defective, but also provide evidence that a feasible, safer alternative design existed at the time the product left the manufacturer's control. This doctrinal shift aligns with the standard articulated in the Restatement (Third) of Torts: Products Liability, which conditions liability for design defects on proof of a reasonable alternative design that would have reduced foreseeable harm [8]. While not uniformly adopted, this requirement has gained traction across U.S. courts, with many treating the absence of such proof as a dispositive failure of the claim.

Importantly, this trend conflicts with emerging doctrines in the EU, which tilt toward easing the plaintiff's burden by presuming causation and defect in certain AI-related harms. Thus, while the European Union is lowering the evidentiary bar to confront the accountability challenges posed by technological opacity through regulatory intervention, the United States is raising it, seeking to preserve established legal standards and aiming to safeguard innovation and economic growth from costly litigation. The message is clear: if AI software functions as a diagnostic or therapeutic agent, it will increasingly be treated as a medical product, with all the liability that entails. The days of sheltering under the "just a service" paradigm are coming to an end. Both the United States and the European Union are converging on the idea that clinical AI should be treated as a product rather than a service, but they are taking different legal paths to get there. The EU is codifying this shift through statutory presumptions and explicit reclassification in its regulatory directives, while the U.S. is approaching the issue more cautiously, through evolving case law and selective application of product liability doctrines.

Regulatory Divergence: FDA Pathways vs. the EU AI Act

We can see now that the regulation of AI in medicine is increasingly bifurcated between two global legal powers: the United States and the European Union. Both recognize the transformative power of AI in healthcare, but their regulatory responses reflect contrasting philosophies. The U.S. approach is incremental and device-centered, grounded in legacy frameworks developed for physical medical technologies. The EU, by contrast, is pursuing a risk-based, system-wide regulatory overhaul grounded in human rights and safety principles.

In the United States, the Food and Drug Administration (FDA) is the primary regulator for software as a medical device (SaMD) products. It offers three primary pathways for market entry:

a. **Premarket Approval (PMA):** The most rigorous pathway, reserved for high-risk (Class III) devices. PMA requires clinical trials and extensive validation, and it confers strong federal preemption protection against state law tort claims under 21 U.S.C. § 360k.

b. **510(k) Clearance:** A faster pathway for devices shown to be "substantially equivalent" to existing products. It is the most common route for AI tools.

c. **De Novo Classification:** For novel, low-to-moderate risk devices without a predicate. Often used for first-in-class AI applications [31].

According to recent FDA analysis, over 90% of AI-driven tools in clinical use today have entered the market through the 510(k) or De Novo pathways rather than PMA [32]. These pathways are less burdensome, but they do not provide strong legal shields. As a result, state-law product liability claims (especially failure to warn claims about AI tools in clinical use) remain viable in most U.S. jurisdictions.

Again, we have only legal precedent from non-AI litigation to consider in evaluating this evolving framework. In Nevolas v. Boston Scientific [33], a patient brought claims against a PMA-approved spinal cord stimulator, alleging design defects, overheating, and inadequate warnings. The court dismissed the case on preemption grounds, holding that the claims were barred under 21 U.S.C. § 360k, which protects manufacturers from state-law requirements that differ from or add to federal PMA conditions. However, the ruling emphasized that preemption is not absolute. Plaintiffs may proceed under a "parallel claim" theory if they can show that the device violated a specific FDA regulation or PMA requirement. In Nevolas

[33], the plaintiff failed to identify any such violation, illustrating how premarket approval can provide a powerful shield, but one that is not impenetrable [33]. The following table summarizes key points of divergence between U.S. and EU liability frameworks for healthcare AI, highlighting differences in regulatory philosophy, liability allocation, and evidentiary standards (Table 1).

Table 1: The following table summarizes key points of divergence between U.S. and EU liability frameworks for healthcare AI, highlighting differences in regulatory philosophy, liability allocation, and evidentiary standards.

Issue	United States	European Union
Regulatory Body	U.S. Food and Drug Administration (FDA) regulates Software as a Medical Device (SaMD).	European Parliament and Council of the EU adopt binding laws; European Commission drafts proposals.
Primary Frameworks	FDA device pathways: Premarket Approval (PMA), 510(k), De Novo.	AI Act (Regulation [EU] 2024/1689); Product Liability Directive (Directive [EU] 2024/2853).
Scope of AI Regulation	Incremental, device centered. AI treated as medical device/software depending on risk classification.	Risk-based, system-wide. Healthcare AI is automatically "high-risk" under the AI Act.
Standard of Care	Defined by state tort law, expert testimony, and evolving clinical guidelines.	Anchored in EU-wide statutory requirements for safety, transparency, and human oversight.
Explainability	No federal requirement: black-box systems may be marketed if performance is validated.	AI Act mandates documentation and context-appropriate explainability for high-risk systems.
Liability Model	Negligence and product liability doctrines; strong preemption defense for PMA devices.	Strict liability under PLD includes software; presumptions of defect and causation for AI.
Burden of Proof	Plaintiff must establish defect or negligence and causation.	Rebuttable presumption of causation/defect if documentation missing or opacity prevents proof.
Preemption	PMA-approved devices shielded from most state-law claims (21 U.S.C. § 360k).	No equivalent: national courts apply harmonized EU liability standards directly.
Compensation Mechanisms	Litigation-driven; no AI-specific no-fault schemes (proposals exist).	Consumer-protective liability regime; emphasis on shifting risk upstream to developers/vendors.
Philosophical Approach	Favors innovation flexibility and post-market adjudication.	Prioritizes precaution, patient safety, and proactive oversight.

Note: The European Commission drafts proposals, but binding Regulations and Directives are adopted by the European Parliament and Council of the European Union. The AI Act and revised Product Liability Directive therefore carry full legal force, while the earlier AI Liability Directive proposal did not.

In the European Union, it is important to distinguish between the institutions: the European Commission drafts and proposes legislation, while binding Regulations and Directives are formally adopted by the European Parliament and Council of the European Union. Accordingly, commission proposals such as the withdrawn AI Liability Directive remain advisory until enacted, whereas the AI Act (Regulation [EU] 2024/1689) and the revised Product Liability Directive (Directive [EU] 2024/2853) carry the force of law

In contrast to the United States, the European Union has adopted a fundamentally different regulatory approach through the AI Act (Regulation [EU] 2024/1689), the revised Product Liability Directive (Directive [EU] 2024/2853), and the proposed AI Liability Directive (AILD). Under the AI Act, all healthcare-related AI systems are automatically classified as "high-risk" and must comply with strict requirements for transparency, human oversight, and ongoing performance monitoring. Standalone software is explicitly treated as a product, subject to conformity assessments. The AILD if adopted would have introduced a rebuttable presumption of causality: if harm results from a high-risk AI system and the developer fails to provide adequate documentation of development and validation, courts may presume the AI caused the harm, shifting the burden of proof to the provider or developer [13,34].

This approach is a sharp departure from the U.S. model, where plaintiffs must still establish direct causation, even when the harm arises from black box systems. As the 2020 European Parliament report warned, "AI opacity will function as a structural barrier to redress" unless reversed by policy intervention [12]. The EU also mandates a harmonized documentation framework, including a post-market surveillance plan and clearly assigned accountability. Under Article 3(3) of the AI Act, each high-risk system must designate a "provider" defined as "a natural or legal person, public authority, agency or other body that develops an AI system or a general-purpose AI model or that has an AI system or a general-purpose AI model developed and places it on the market or puts the AI system into service under its own name or trademark, whether for payment or free of charge" [34,35].

From a clinician or hospital perspective, this means that under EU law, the burden is structurally redistributed upstream, toward developers and vendors. In the U.S., where litigation risk remains fragmented and largely driven by state law tort doctrines, clinicians and hospitals may bear more residual liability. As outlined in the 2025 Stanford Technology Law Review, this divergence has profound consequences for global developers: "Companies building medical AI must design systems not just for clinical safety, but for jurisdictional survivability. What passes in Boston may fail in Berlin" [18]. Ultimately, neither system is without trade-offs. The U.S. offers regulatory flexibility but lacks cohesive legal protection. The EU provides structured guardrails and patient-centric presumptions,

but at the cost of higher up-front compliance and development costs. For global developers, the challenge is to navigate both landscapes without compromising clinical integrity or legal risk.

Toward a New Legal Framework for AI Accountability

As AI systems become embedded in clinical workflows, existing liability principles must evolve beyond frameworks designed for static tools and human error. Traditional tort doctrines are ill-equipped to handle adaptive algorithms that operate without clear attribution or auditability. Instead of relying solely on reactive litigation, policymakers must develop a forward-looking legal infrastructure. This framework should allocate responsibility across the full ecosystem, including developers, institutions, physicians and regulators, to ensure accountability at every stage. It must also recognize that harm can result from complex system interactions, even when no single party or failure can be clearly identified.

Several legal frameworks have been proposed to bridge the accountability gap in healthcare AI. Each approaches risk allocation differently and reflects competing priorities such as protecting patients versus supporting industry, encouraging innovation versus maintaining stability, and balancing fairness with feasibility.

Enterprise liability

Under an enterprise liability model, legal responsibility for AI-related harm is assigned to the healthcare institution that deploys the system-regardless of whether the harm arose from clinician error, software design, or integration failure. This simplifies litigation by identifying a single, well-resourced defendant and encourages internal risk governance. Hospitals already assume vicarious liability for the actions of their staff. Enterprise liability builds on this foundation by extending institutional responsibility to the algorithms they deploy. When an AI system operates under the hospital's authority, the institution may justly be held accountable for its consequences. Chew et al. [22] We've seen this logic applied in other areas: hospitals are liable for credentialing decisions, medical device maintenance, and staffing. Extending that accountability to AI oversight would incentivize robust deployment protocols, better vendor evaluation, and post-market surveillance. One longstanding criticism of enterprise liability is that it effectively enshrines the "deep-pocket" theory, where institutional defendants-typically better resourced than individual providers-are held morally or financially responsible, regardless of fault [36,37].

No-fault compensation funds

Another approach is to establish no-fault compensation schemes, modeled on the National Vaccine Injury Compensation Program. In this system, patients harmed by AI would receive compensation without having to prove negligence or defect. These funds could be financed by levies on AI vendors, insurers, or healthcare institutions.

This approach emphasizes access to justice and patient trust, particularly in cases involving systemic opacity or "black box" decision-making. As the 2024 NEJM article framed it: "Where proof breaks down, justice must not" [38]. No-fault mechanisms would lower litigation burdens, reduce defensive posturing, and restore public confidence in AI systems that are difficult-if not impossible-for individuals to challenge through traditional tort theories of liability. While no-fault compensation systems provide a streamlined alternative to traditional litigation, they come with notable limitations. Critics argue that these programs often restrict coverage to narrowly defined injuries, excluding many patients who suffered legitimate harm. Despite their goal of promoting transparency and learning, there is little evidence that no-fault systems lead to meaningful improvements in patient safety or clinical practices. Others caution that by encouraging a higher volume of claims, these systems may raise overall costs and administrative burdens without delivering greater accountability or deterrence [39,40].

Mandatory AI liability insurance

A market-based model would require developers and/or deployers of high-risk AI systems to carry specialized liability insurance. Premiums would reflect the risk profile of the product, and insurers would effectively become third-party regulators, demanding documentation, audit trails, and safety features before underwriting policies. The 2025 Science policy commentary put it plainly: "The point of AI insurance is not just payout-it's prevention. Carriers demand audit trails, error logs, and fail-safes. Liability premiums become the cost of opacity" [26].

Mandatory AI liability insurance, however, faces scrutiny for several valid concerns. First, small developers may be disproportionately burdened by high premiums or limited access to coverage. Second, there's a real risk of moral hazard, where insurance might disincentivize rigorous safety practices. Third, broad or poorly defined policies may lead to coverage gaps, especially given the novelty and complexity of AI risks. Finally, insurers historically struggle to understand and price AI-related risk accurately, making underwriting uncertain and potentially leaving exposures underinsured or overly costly to insure [41,42].

Transparency and documentation mandates

Regardless of the liability model adopted, transparency remains foundational. Without comprehensive documentation-covering training datasets, known limitations, update histories, and evidence of behavioral drift-any system of accountability, whether legal or clinical, collapses. The EU AI Act embeds these documentation requirements directly into high-risk system mandates, demanding data governance, technical files, and post-market surveillance [43]. In parallel, the FDA's January 2025 draft guidance for AI-enabled medical devices emphasizes lifecycle documentation, including model descriptions, data management plans, validation protocols, and performance monitoring strategies [44,45].

Unfortunately, translating these formal mandates into consistent practice remains a mixed landscape, often leaving implementation dependent on vendor voluntarism rather than uniform compliance. As a recent BMC Medical Ethics review emphasizes, thorough documentation serves not as a perfunctory requirement but as

an essential bridge between opaque algorithmic processes and accountable human oversight [21]. Clinicians and institutions must demand transparency up front. If a vendor cannot or will not disclose how a system was trained or how it updates, that is not a proprietary strength. It is a legal vulnerability.

Shared accountability and dynamic consent

Legal scholars are increasingly exploring frameworks of distributed responsibility, where liability is proportionally shared among stakeholders based on real-time attribution. Under such models, clinicians would explicitly document AI involvement in clinical notes, patients would receive dynamic consent that discloses AI's role in care decisions, and developers would be obligated to maintain performance guarantees linked to monitored system outputs. Implementing this approach would require significant infrastructural reform: integration of AI audit trails into electronic health records, standardized clinician training protocols, and accessible patient education tools. Nevertheless, this paradigm shift reflects a broader evolution in medical law-away from retrospective blame assignment and toward proactive system design, where accountability is embedded across every layer of clinical interaction [46].

Conclusion: Building Accountability in the Age of Clinical AI

The integration of artificial intelligence into clinical practice presents both unprecedented opportunities and profound legal challenges. While these tools promise improved accuracy and efficiency, they introduce risks that traditional tort law was not designed to manage. Existing legal liability frameworks assume human agency, fixed products, and traceable causation. AI systems, by contrast, evolve post-deployment, operate autonomously, and often obscure attribution. Addressing these differences requires a shift from reactive litigation to proactive governance. Transparency, auditability, and institutional accountability must be built into every stage of development, deployment, and clinical use. Regulatory models are diverging. The European Union favors pre-market controls, documentation mandates, and presumptions of liability. The United States relies more heavily on post-market adjudication and innovation flexibility. Both systems demand greater legal clarity and institutional readiness.

Clinicians must document AI involvement in medical records, maintain independent oversight, and understand tool limitations. Hospitals must establish governance protocols for validation, monitoring, and incident response. Developers must treat clinical AI as a high-liability product, with version control, performance tracking, and accessible documentation. Regulators must balance innovation with patient protection, ensuring both speed and safety. Justice in this new landscape will not come from isolating fault but from designing systems that anticipate failure, distribute responsibility, and protect human dignity. Accountability must be operationalized if AI is to fulfill its promise in healthcare without compromising trust or safety.

References

1. Henry T (2025) 2 in 3 physicians are using health AI-up 78% from 2023. American Medical Association (AMA), Chicago, Illinois, USA.
2. Dahdah R (2024) Microsoft makes the promise of AI in healthcare real through new collaborations with healthcare organizations and partners.
3. Elsevier Limited (2025) Elsevier's clinician of the future 2025 survey: Clinicians' AI usage and optimism grows despite concerns around trust and reliability.
4. North M (2025) 7 Ways AI Is transforming healthcare, Health and Healthcare Systems, World Economic Forum (WEF), Cologny, Switzerland.
5. Dobbs DB, Hayden PC, Bublick E (2011) The law of torts. (2nd edn), West Academic Publishing, Minnesota, USA.
6. (1965) Restatement (Second) of torts, American Law Institute, Pennsylvania, USA.
7. Hall MA, Bobinski MA, Orentlicher D (2013) Medical liability and treatment relationships. (3rd edn), Aspen Publishers, USA.
8. (1998) Restatement (Third) of torts: Products liability, American Law Institute, Pennsylvania, USA.
9. Keeton P (1984) Prosser and Keeton on torts. (5th edn), West Publishing Company, Minnesota, USA.
10. Verghese A, Shah NH, Harrington RA (2023) Revisiting the EMR revolution: A cautionary tale for AI. JAMA 330(10): 932-934.
11. Nouis SC, Uren V, Jariwala S (2025) Evaluating accountability, transparency, and bias in AI-assisted healthcare decision-making: A qualitative study of healthcare professionals' perspectives in the UK. BMC Medical Ethics 26(1): 89.
12. European Parliamentary Research Service (2020) Artificial intelligence in healthcare and liability: Policy approaches (No. PE 641.547), European Parliament, Strasbourg, France.
13. European Parliamentary Research Service (2023) Artificial intelligence liability directive (No. PE 739.342), European Parliament, Strasbourg, France.
14. European Parliament and Council of the European Union (2024) Directive (EU) 2024/2853 of the European parliament and of the council on liability for defective products repealing council directive 85/374/EEC, European Union, Brussels, Belgium.
15. Reed Smith LLP (2025) AI liability directive and PLD revision: How the EU is Reshaping Product Law.
16. European Parliament and Council of the European Union (2024) Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 on harmonized rules on artificial intelligence (Artificial Intelligence Act). Official Journal of the European Union L138: 1-92.
17. Mello MM, Guha N (2024) Understanding liability risk from using health care artificial intelligence tools. New England Journal of Medicine 390(3): 271-278.
18. Dhar M, Walsh E (2025) Opacity and accountability in medical AI: Reconstructing harm in black box systems. Stanford Technology Law Review 28(1): 45-78.
19. (2018) Skounakis v. Sotillo.
20. Kersten MS (2022) Professionally responsible artificial intelligence. Northwestern University Law Review 117(1): 163-220.
21. Corfmat M, Martineau JT, Régis C (2025) High-reward, high-risk technologies? An ethical and legal account of AI development in healthcare. BMC Medical Ethics 26(1): 4.

22. Chew K, Snyder K, Pert C (2025) How physicians might get in trouble using AI (or not using AI). Missouri Medicine 122(3): 169-172.

23. (2019) Proposed regulatory framework for modifications to Artificial Intelligence/Machine Learning (AI/ML)-Based Software as a Medical Device (SaMD). US Food and Drug Administration, Maryland, USA.

24. KFF Health News (2025) Health care AI, intended to save money, turns out to need expensive human support. KFF Health News.

25. Regulation (EU) 2024/1689-Article 6 and Annex III (2024).

26. Topol EJ, Tufekci Z (2025) Evolving accountability in translational AI. Science 378(6622): 994-997.

27. (2017) Taylor v. Intuitive Surgical Inc, 389 P.3d 517.

28. Reardon R (2024) The erosion of the learned intermediary doctrine in the age of clinical AI. DePaul Law Review 73(3): 627-661.

29. (2022) Lowe v Cerner Health Servs.

30. European Commission (2023) Proposal for a directive on adapting non-contractual civil liability rules to artificial intelligence (AI Liability Directive) and revision of the Product Liability Directive. European Commission.

31. US Food and Drug Administration (2025b) Artificial intelligence in software as a medical device. Software as a Medical Device (SaMD).

32. US Food and Drug Administration (2024) Artificial intelligence and machine learning (AI/ML)-enabled medical devices. Software as a Medical Device (SaMD).

33. Western District of Oklahoma (2016) Nevolas v. Boston scientific corporation, No. 5:2015cv00894-Document 37 (W.D. Okla. 2016). JUSTIA US Law.

34. European Parliament and Council of the European Union (2024) General provisions, Article 3(3) Definition of "Provider," AI Act (Regulation (EU) 2024/1689).

35. EYREACT (2025) Who is a provider under the AI Act?

36. MacCoun RJ (1996) Differential treatment of corporate defendants by juries: An examination of the "Deep-Pockets" hypothesis. Law & Society Review 30(1): 121-161.

37. Morris C (1961) Enterprise liability and the acts of God. The Yale Law Journal 70(4): 554-582.

38. Cohen IG, Spector-Bagdady K (2024) Paging Dr. Robot. New England Journal of Medicine 390(12): 1123-1126.

39. Gaine WJ (2003) No-fault compensation systems. British Medical Journal 326(7397): 997-998.

40. Studdert DM, Brennan TA (2001) No-fault compensation for medical injuries: The prospect for error prevention. Journal of the American Medical Association 286(2): 217-223.

41. Deloitte Insights (2024) Risk insurance for AI coverage.

42. Row S (2025) AI and insurance-The awkward early days. Stoel Rives LLP.

43. Aboy M, Minssen T, Vayena E (2024) Navigating the EU AI act: Implications for regulated digital medical products. NPJ Digital Medicine 7(1): 237.

44. US Food and Drug Administration (2025) Draft guidance: Artificial intelligence-enabled device software functions-lifecycle documentation.

45. US Food and Drug Administration (2025) FDA issues comprehensive draft guidance for developers of artificial intelligence-enabled medical devices.

46. Wang W, Wang Y, Chen L, Ma R, Zhang M (2024) Justice at the forefront: Cultivating felt accountability towards Artificial Intelligence among healthcare professionals. Social Science & Medicine 347: 116717.

