Indonesian Health Justice Review
ISSN: XXXX-XXXXX (Online)
Vol. 3, No. 1, 2026, pp. 99-118
Published online since March 25, 2026

Malpractice or System Failure? Reconstructing Medical Liability for AI-Assisted Diagnostic Errors in Indonesia

Putri Fahima*
Universitas Diponegoro, Semarang, Indonesia
putrifahima.mhsundip@gmail.com

DOI: https://doi.org/10.65815/r46feh64

Submitted: 30-08-2025	Revised: 29-11-2025, 27-02-2026	Accepted: 10-03-2026

*Corresponding Author

Abstract

AI-assisted diagnostic technologies challenge conventional concepts of medical malpractice because patient harm may result from interactions between human judgment and algorithmic systems. When a physician relies on an incorrect AI-generated recommendation, determining whether the resulting harm constitutes medical negligence, technological failure, or both becomes legally complex. This article investigates the appropriate legal characterization and allocation of liability for AI-assisted diagnostic errors in Indonesia. Through normative legal analysis, the study examines the standards of professional medical care, negligence, institutional responsibility, product-related liability, and emerging principles of AI accountability. The research finds that applying conventional malpractice doctrines exclusively to physicians may inadequately address errors caused by defective algorithms, biased training datasets, inadequate validation, or insufficient warnings from technology providers. Conversely, imposing strict liability on developers for every clinical error may undermine innovation and fail to recognize the physician's independent professional judgment. The article proposes a layered accountability model distinguishing clinical negligence, technological defects, inadequate system governance, and shared causation. Liability should correspond to each actor's ability to prevent, identify, and mitigate the relevant risk. This approach would provide more equitable remedies for patients while establishing clearer responsibilities for physicians, hospitals, and AI developers in Indonesia's emerging digital healthcare ecosystem.

[*Teknologi diagnostik berbantuan AI menantang konsep konvensional tentang malpraktik medis karena bahaya yang diderita pasien dapat terjadi akibat interaksi antara penilaian manusia dan sistem algoritmik. Ketika seorang dokter bergantung pada rekomendasi yang salah yang dihasilkan oleh AI, menentukan apakah bahaya yang dihasilkan merupakan kelalaian medis, kegagalan teknologi, atau keduanya menjadi kompleks secara hukum.*

Artikel ini menyelidiki karakterisasi hukum yang tepat dan alokasi tanggung jawab atas kesalahan diagnostik berbantuan AI di Indonesia. Melalui analisis hukum normatif, studi ini meneliti standar perawatan medis profesional, kelalaian, tanggung jawab institusional, tanggung jawab terkait produk, dan prinsip-prinsip akuntabilitas AI yang muncul. Penelitian ini menemukan bahwa penerapan doktrin malpraktik konvensional secara eksklusif kepada dokter mungkin tidak cukup untuk mengatasi kesalahan yang disebabkan oleh algoritma yang cacat, kumpulan data pelatihan yang bias, validasi yang tidak memadai, atau peringatan yang tidak cukup dari penyedia teknologi. Sebaliknya, penerapan tanggung jawab ketat pada pengembang untuk setiap kesalahan klinis dapat merusak inovasi dan gagal mengakui penilaian profesional independen dokter. Artikel ini mengusulkan model akuntabilitas berlapis yang membedakan kelalaian klinis, cacat teknologi, tata kelola sistem yang tidak memadai, dan penyebab bersama. Tanggung jawab hukum harus sesuai dengan kemampuan masing-masing pihak untuk mencegah, mengidentifikasi, dan mengurangi risiko yang relevan. Pendekatan ini akan memberikan solusi yang lebih adil bagi pasien sekaligus menetapkan tanggung jawab yang lebih jelas bagi dokter, rumah sakit, dan pengembang AI dalam ekosistem layanan kesehatan digital yang sedang berkembang di Indonesia.]

Keywords: Medical malpractice, AI-assisted diagnosis, negligence, liability, healthcare technology

Introduction

Artificial intelligence is increasingly changing the conditions under which medical decisions are produced. Diagnostic algorithms can analyze medical images, identify patterns in laboratory results, estimate clinical risk, prioritize patients, and generate recommendations that would previously have depended almost entirely upon human clinical judgment. The legal significance of this transformation becomes apparent when the technology fails. If an AI system incorrectly classifies a radiological image, for example, and a physician accepts the recommendation without detecting the error, the resulting injury cannot easily be characterized as either purely medical malpractice or purely technological failure. The harm emerges from an interaction between software, clinical judgment, institutional deployment, and the information environment in which the system operates. Contemporary scholarship consequently recognizes that existing liability frameworks may be inadequate when applied to AI-mediated healthcare without modification (Parikh et al. 2021; Smith and Fotheringham 2020; Cestonaro et al. 2023).

The conventional law of medical malpractice is generally organized around the conduct of the healthcare professional. The central inquiry is whether the

physician owed a duty to the patient, breached the applicable professional standard, and thereby caused compensable harm. This framework has considerable normative strength because physicians exercise professional judgment over matters directly affecting patients' health and bodily integrity. It also prevents clinicians from avoiding responsibility merely by attributing an unfavorable outcome to the tools they have chosen to use. Yet AI creates a new problem because the relevant "tool" may itself contain defects that are not reasonably detectable by the physician. A diagnostic model may have been trained on unrepresentative data, validated in a different patient population, deployed outside its intended clinical environment, inadequately updated, or affected by an error that cannot reasonably be identified from the output alone (Gianfrancesco et al. 2018; Kelly et al. 2019; Cestonaro et al. 2023).

The opposite approach is equally problematic. If every injury associated with AI is attributed to the developer or manufacturer, the law would disregard the physician's continuing obligation to exercise independent professional judgment. AI-assisted diagnosis does not necessarily transfer the clinical decision from the physician to the machine. In many applications, the algorithm provides a recommendation while the physician retains formal authority to accept, reject, or contextualize it. The legal challenge is therefore not to determine whether AI or the physician is "the decision-maker" in an abstract sense. It is to determine which actor controlled the relevant risk and which actor was reasonably capable of preventing the harm.

This problem is particularly significant in Indonesia because the country's health-law framework has undergone substantial restructuring. Law No. 17 of 2023 concerning Health repealed and replaced several important statutes, including Law No. 29 of 2004 concerning Medical Practice, Law No. 36 of 2009 concerning Health, and Law No. 44 of 2009 concerning Hospitals. The new framework expressly addresses patient safety, quality control, professional discipline, dispute resolution, health technology, and the responsibilities of healthcare institutions. Government Regulation No. 28 of 2024 subsequently provided implementing rules for Law No. 17 of 2023, including provisions concerning health services, healthcare personnel, health facilities, health technology, health information systems, and supervision. Yet neither statute establishes a comprehensive liability regime specifically designed for AI-assisted diagnosis.

The resulting regulatory gap should not be understood simply as an absence of legislation. The more difficult question is how existing legal principles should be reconstructed when clinical harm is produced by a sociotechnical system rather than by an isolated human act. Law No. 17 of 2023 requires medical and healthcare personnel to observe patient safety and establishes quality control responsibilities within healthcare facilities. Article 303 expressly provides that healthcare facilities bear responsibility for quality and cost control, while Article 304 establishes professional-discipline enforcement and Article 305 permits patients or their families whose interests have been harmed by healthcare personnel to submit

complaints. These provisions provide a foundation for distinguishing individual professional negligence from institutional governance failure.

The principal argument of this article is that Indonesia should not approach AI-assisted diagnostic injury through a binary choice between "medical malpractice" and "technology failure." The more appropriate framework is a layered model of accountability. Physician liability should arise where the clinician negligently selects, interprets, supervises, or relies upon an AI system in circumstances where a reasonably competent practitioner should have recognized the risk. Developer or provider responsibility should arise where harm results from defects in design, training, validation, instructions, warnings, cybersecurity, or other aspects within the provider's sphere of control. Institutional liability should arise where a healthcare facility deploys an AI system without adequate validation, monitoring, training, governance, or safeguards. Where several actors materially contribute to the same injury, causation should be analyzed as shared rather than artificially reduced to a single responsible party.

The article employs normative legal research, examining legislation and regulatory materials together with legal and interdisciplinary scholarship on medical negligence, AI governance, algorithmic bias, explainability, and healthcare liability. The analysis is doctrinal and conceptual rather than empirical. It asks how existing Indonesian legal principles can be interpreted and developed to address AI-related harm and what additional regulatory structure would be necessary to make responsibility more predictable.

Existing scholarship has established that AI complicates traditional medical liability because responsibility can be distributed across physicians, hospitals, developers, and other actors. Parikh et al. (2021) argue that conventional malpractice frameworks are inadequate to address the broader AI ecosystem, while Smith and Fotheringham (2020) contend that software developers may owe duties alongside clinicians. Mello and Guha (2024) similarly demonstrate that AI-related liability involves questions concerning software errors, clinical reliance, organizational adoption, and risk management. Cestonaro et al. (2023), through a systematic review, identify malpractice, vicarious liability, product liability, and other possible approaches while emphasizing the absence of a definitive liability model for AI diagnostic errors. These studies, however, largely focus on the United States, Europe, or general comparative theory. The Indonesian problem requires a further step: determining how AI-related responsibility can be reconstructed within Indonesia's current health-law architecture, particularly following Law No. 17 of 2023.

The article therefore addresses three interconnected questions. First, when does reliance on an AI diagnostic system constitute medical negligence under Indonesian law? Second, when should responsibility shift from the physician toward the healthcare institution or technology provider? Third, how should Indonesian law allocate responsibility when an AI-assisted diagnostic error results from multiple interacting causes? The article argues that these questions should be

answered through a risk-based allocation of responsibility rather than through an exclusive physician-centered malpractice model.

Reconstructing Medical Negligence in the AI-Assisted Clinical Environment

Medical malpractice traditionally assumes that the physician is the principal source of clinical judgment and that the relevant standard of care can be assessed by examining professional conduct against the practices expected of a reasonably competent practitioner. AI complicates this assumption because the physician may now make a clinical decision within an environment in which part of the diagnostic reasoning has been generated by an algorithm. The legal standard should therefore not simply ask whether the physician followed or departed from an AI recommendation. It should ask whether the physician used the AI system in a professionally reasonable manner.

This distinction is crucial because following an AI recommendation cannot automatically establish negligence. A physician who accepts an AI-generated recommendation may have acted reasonably if the system was appropriately validated, the recommendation was consistent with the clinical evidence, the physician considered relevant patient-specific information, and no circumstances reasonably indicated that the output was unreliable. Conversely, refusing to follow an AI recommendation should not automatically constitute negligence merely because the algorithm subsequently proves to have been correct. Medical negligence must remain an ex ante inquiry into the reasonableness of professional conduct at the time the decision was made.

The same principle applies to the physician who rejects an AI recommendation. If a validated diagnostic system identifies a clinically significant abnormality and the physician disregards it without adequate clinical justification, the use of AI may become relevant evidence of negligence. The important issue is not whether the physician followed the machine but whether the physician exercised appropriate professional judgment in light of all reasonably available information. Scholarship on AI and malpractice similarly recognizes that both overreliance and inappropriate disregard of AI recommendations can create liability concerns (Price, Gerke, and Cohen 2021; Tobia et al. 2021).

The legal standard must therefore accommodate the possibility that AI changes what constitutes reasonable professional conduct. If an AI system has become an accepted and reliable component of a particular specialty's diagnostic practice, a physician's complete refusal to consider its output may eventually become difficult to reconcile with the professional standard of care. Conversely, if the technology is experimental, insufficiently validated, or known to perform poorly for the relevant patient population, reliance upon it may itself constitute a departure from reasonable practice.

This does not mean that AI should automatically become the new legal standard of care. Such an approach would create a dangerous circularity in which technology providers effectively determine professional obligations by marketing their systems as superior to human clinicians. The standard of care must remain a professional and contextual legal standard. AI performance may constitute evidence relevant to that standard, but it should not automatically dictate it.

A further difficulty arises from automation bias. Physicians may place excessive confidence in algorithmic recommendations because the system appears objective, data-driven, or statistically sophisticated. The risk is particularly acute when the algorithm's output is presented with a level of confidence that exceeds the reliability of the underlying model. The European Union's AI framework recognizes this problem by requiring human oversight for high-risk AI systems and expressly identifying the risk of over-reliance on automated outputs. Human overseers must be capable of understanding system limitations, interpreting outputs, and overriding or disregarding the system when necessaryAlthough the EU framework is not binding in Indonesia, its approach illustrates a useful regulatory principle: human involvement is meaningful only when the human decision-maker possesses sufficient competence, authority, and practical capacity to challenge the system.

The same principle should inform Indonesian medical liability. A physician should not be able to defend against malpractice merely by stating that “the AI told me so.” At the same time, a physician should not automatically become liable merely because an AI system produced an incorrect output. Liability should depend upon whether the physician's reliance was reasonable given the system's intended use, known limitations, patient circumstances, available alternatives, and the information reasonably accessible to the clinician.

Law No. 17 of 2023 strengthens this interpretation because patient safety is expressly embedded within healthcare practice. The statute requires medical and healthcare personnel to observe patient safety, while healthcare facilities bear responsibility for quality and cost control. AI-assisted diagnosis should consequently be treated as part of the clinical environment governed by the general obligation of safe and quality healthcare rather than as a separate technological domain immune from professional standards. The decisive question should therefore be whether the physician exercised reasonable clinical judgment in the presence of AI, rather than whether the physician exercised judgment instead of AI. This formulation preserves the traditional function of malpractice law while adapting its factual inquiry to technological change.

When an AI Error Becomes a System Failure

The most significant weakness of a purely physician-centered malpractice model is that some AI errors are not reasonably detectable by clinicians. A diagnostic algorithm may produce a systematically biased result because its training dataset

inadequately represented a particular population. It may fail because the deployment environment differs from the environment in which the model was validated. It may malfunction following a software update. It may be affected by data-quality problems, model drift, or an undisclosed limitation that is known to the developer but not to the physician. In such circumstances, characterizing the entire injury as physician negligence obscures the source of the relevant risk.

Algorithmic bias illustrates the point particularly clearly. Bias may emerge during data collection, annotation, model development, validation, or deployment (Gianfrancesco et al. 2018). Obermeyer et al. (2019), for example, demonstrated how an algorithm used in healthcare management could generate significant racial disparities because the variable selected as a proxy for health need reflected unequal patterns of healthcare expenditure. The lesson for liability law is not simply that algorithms can be biased. It is that the source of the error may exist upstream from the physician's interaction with the system.

If a physician receives a recommendation generated by a commercially developed system and has no reasonable means of identifying that the underlying model performs poorly for the patient's demographic group, attributing the resulting harm exclusively to the physician may be normatively inappropriate. The physician may have failed to diagnose the condition, but the causal explanation for that failure may lie partly in the design and deployment of the technology.

The concept of system failure therefore provides a necessary complement to malpractice. A system failure occurs when the injury results from conditions in the AI ecosystem that fall outside the physician's reasonable capacity to control, including defective design, inadequate training data, inappropriate validation, misleading performance claims, inadequate warnings, insufficient monitoring, or inappropriate deployment. The concept should not be treated as a way of eliminating professional responsibility. Instead, it identifies circumstances in which responsibility exists beyond the physician.

This approach is supported by contemporary scholarship. Parikh et al. (2021) argue that the AI ecosystem includes multiple stakeholders and that conventional malpractice frameworks may not adequately incentivize safe clinical implementation. Smith and Fotheringham (2020) similarly argue that software developers should not necessarily remain outside the scope of patient-protective duties merely because physicians interact directly with patients. Cestonaro et al. (2023) find that existing literature considers physician malpractice, hospital responsibility, product liability, and other models because AI-related harm may arise at different points in the technological chain.

The implication for Indonesia is that the healthcare institution should have an independent governance responsibility. A hospital that purchases an AI diagnostic system cannot reasonably treat the vendor's marketing materials as a complete substitute for institutional due diligence. The institution should determine whether the system is appropriate for its clinical environment, whether its personnel are sufficiently trained, whether its performance is monitored, and whether procedures

exist for detecting and responding to errors. These responsibilities are particularly compatible with Article 303 of Law No. 17 of 2023, which places quality-control responsibilities on healthcare facilities.

The institutional dimension is also reinforced by Government Regulation No. 28 of 2024, which implements Law No. 17 of 2023 across healthcare delivery, health facilities, health information systems, health technology, and supervision. The regulation does not establish a complete AI liability doctrine, but its broader governance architecture supports the proposition that technological systems used in healthcare should be embedded within institutional quality and safety mechanisms.

This means that a hospital could potentially be responsible even where the physician acted reasonably. Suppose a hospital deploys an AI diagnostic tool without testing whether its performance is appropriate for the hospital's patient population. The physician then uses the system in accordance with institutional protocols, but the algorithm produces a systematic error that results in serious harm. It would be artificial to characterize the case exclusively as physician malpractice. The institution's failure to validate or monitor the system may constitute an independent causal contribution.

The same reasoning applies where a hospital knows that physicians routinely over-rely on the AI system but fails to establish meaningful oversight. If the institution provides no training concerning automation bias, no procedure for overriding AI recommendations, and no monitoring of diagnostic discrepancies, its responsibility is not merely derivative of the physician's conduct. It arises from an independent failure of governance.

This approach also prevents a problematic phenomenon in which hospitals effectively outsource safety responsibilities to technology providers. The fact that an AI system has been developed by a reputable company or has obtained regulatory authorization does not eliminate the institution's responsibility to use it appropriately within its own clinical environment. Validation is context-dependent. A system can perform adequately under one set of conditions and inadequately under another.

Developer Liability and the Limits of Product-Based Responsibility

If physician malpractice is insufficient to capture AI-related harm, the next question is whether the technology developer should bear responsibility. Product-related liability provides an intuitive starting point because AI systems may contain defects that cause injury. Yet conventional product-liability concepts are not perfectly suited to adaptive software and machine-learning systems.

A traditional defective-product analysis assumes that the product has identifiable characteristics at the time it enters the market and that the defect can be connected to the manufacturer's conduct. AI can be more complicated. The system may be updated after deployment, learn from new data, operate differently

in different environments, or generate outputs that depend upon interactions among software, user behavior, and patient data. The relevant defect may therefore not be contained in a single line of code or identifiable at a single point in time.

The difficulty should not, however, lead to immunity for developers. A technology provider that designs a diagnostic system for clinical use possesses information and technical capacities that physicians and hospitals ordinarily do not. The developer is therefore better positioned to identify certain risks associated with architecture, training data, model validation, cybersecurity, known failure modes, and intended-use limitations. Liability rules should reflect this asymmetry of knowledge and control.

This is consistent with a broader principle of risk allocation: responsibility should follow the actor's capacity to prevent or mitigate the relevant risk. If a developer knows that a diagnostic model performs poorly for a particular population and fails to communicate that limitation, responsibility should not be shifted entirely to the physician who relies upon the system without access to that information. Similarly, if a software provider markets a system for a clinical purpose outside the context in which it was validated, the provider's contribution to resulting harm becomes legally significant.

The principle is visible in international AI governance. The World Medical Association's 2025 statement emphasizes that AI in medicine must remain subject to human accountability and that responsibility is shared across physicians, healthcare organizations, developers, regulators, and other stakeholders. It also recognizes that patients should receive understandable information about AI limitations and potential errors. The statement does not establish Indonesian law, but it reflects an increasingly influential conception of AI responsibility as distributed across the technological ecosystem.

The National Institute of Standards and Technology similarly treats trustworthy AI as involving multiple characteristics, including validity and reliability, safety, security, accountability, transparency, explainability, privacy, and fairness (Tabassi 2023). These characteristics have direct relevance to liability because a system's failure in any one dimension may contribute to patient injury. A model can be statistically accurate while still being unsafe in a particular clinical setting. It can be technically functional while being inadequately transparent. It can be well designed while being improperly deployed.

The legal problem is therefore not simply whether an AI system is “defective.” It is whether a particular failure falls within the sphere of responsibility that the developer could reasonably control. This requires distinguishing between design defects, manufacturing or coding defects, inadequate warnings or instructions, inadequate validation, and post-deployment failures.

The warning dimension is especially important. Physicians cannot be expected to independently discover every technical limitation of a complex model. If the provider knows that the system has significant limitations but communicates only generalized performance statistics, the physician may be misled into believing

that the system is appropriate for a broader population or purpose than is actually the case. In such circumstances, inadequate warnings can become an independent basis for responsibility.

The same reasoning supports greater transparency concerning performance. The 2024 guiding principles developed by the FDA, Health Canada, and the MHRA emphasize communication of intended use, development, performance, limitations, and the human-AI relationship for machine-learning-enabled medical devices. Transparency is treated not merely as an informational preference but as a condition relevant to safe clinical use. This approach is consistent with the idea that the actor possessing the greatest technical knowledge should bear corresponding duties to communicate material limitations.

Nevertheless, imposing strict developer liability for every AI-related injury would be equally problematic. An AI system may produce an incorrect recommendation despite being properly designed, appropriately validated, and used within its intended environment. Clinical decisions involve uncertainty, and not every incorrect prediction is evidence of a defect. If developers were automatically liable whenever an algorithmic recommendation contributed to injury, the legal system would effectively guarantee the accuracy of probabilistic technologies that are inherently incapable of eliminating diagnostic uncertainty.

The appropriate approach is therefore neither complete immunity nor automatic strict liability. Developer responsibility should be tied to controllable technological failures, including defective design, inadequate validation, known limitations that were not adequately disclosed, foreseeable misuse that was not appropriately addressed, or failures to monitor and correct material defects. Where the system simply produces a clinically incorrect prediction despite satisfying reasonable standards of design and validation, the existence of an adverse outcome alone should not establish developer liability.

Shared Causation and the Architecture of Layered Accountability

The most difficult cases will not fit neatly into a single category. An AI diagnostic error may simultaneously involve a physician's excessive reliance, a hospital's inadequate training, and a developer's failure to disclose a known limitation. The legal system must therefore avoid the assumption that every case has one identifiable wrongdoer. AI-mediated healthcare frequently produces what may be described as distributed causation, in which several decisions and conditions interact to produce the final injury.

Distributed causation does not mean that responsibility becomes impossible to determine. It requires the legal inquiry to be divided into causally and normatively meaningful questions. The first question is whether the AI system itself contained a relevant defect or limitation. The second is whether the healthcare institution appropriately selected, validated, deployed, and monitored the system. The third is whether the physician used and interpreted the system in accordance

with reasonable professional standards. The fourth is whether the patient's injury was sufficiently connected to these failures to establish legal causation.

This structure is preferable to a model in which the physician automatically becomes the residual defendant whenever an AI-assisted diagnosis proves wrong. Such a residual model would create distorted incentives. Physicians might avoid beneficial AI tools because using them increases their exposure to malpractice claims even when the technology is appropriately deployed. Developers, meanwhile, might have insufficient incentives to improve system safety if responsibility is consistently shifted downstream to clinicians.

The opposite distortion is also possible. If physicians and hospitals assume that technology providers bear all AI-related responsibility, clinicians may become excessively dependent on automated recommendations. The result could be precisely the kind of automation bias that responsible AI governance seeks to prevent. Liability must therefore operate as an incentive structure for all actors, not merely as a mechanism for compensating patients after harm occurs.

Parikh et al. (2021) identify this incentive problem directly, arguing that liability rules should encourage safe implementation without unnecessarily suppressing innovation. Mello and Guha (2024) similarly emphasize the need to understand liability risks arising from healthcare AI adoption. Their analysis suggests that healthcare organizations and physicians must consider how software-related errors interact with existing malpractice principles rather than assuming that AI creates an entirely separate legal category.

Indonesian law already contains elements that support this layered approach. Law No. 17 of 2023 distinguishes professional discipline from broader dispute resolution. Patients or families whose interests are harmed by medical or healthcare personnel may submit complaints concerning professional conduct, while disputes arising from alleged professional errors are subject to alternative dispute resolution mechanisms before litigation. The statutory structure therefore recognizes that professional responsibility has institutional procedures of its own. AI-related cases should build upon, rather than bypass, this architecture.

The existence of professional discipline, however, should not be confused with civil liability. A physician may commit a disciplinary violation without causing compensable damage, while a patient may suffer compensable injury even when no disciplinary violation is established. AI-related disputes will require courts and other decision-makers to distinguish these questions carefully. A system failure may also exist even where the physician's professional conduct was reasonable.

The causal analysis should consequently focus on the risk contribution of each actor. If a physician negligently ignores obvious clinical inconsistencies in an AI recommendation, the physician's conduct may constitute a substantial causal contribution. If the physician could not reasonably have identified the algorithmic error because the developer concealed a known limitation, the developer's contribution becomes more significant. If the hospital failed to conduct basic validation before deployment, institutional responsibility may coexist with both.

The concept of comparative responsibility may provide a useful analytical analogy, although Indonesian courts should not mechanically import foreign tort doctrines. The broader principle is that responsibility should correspond to each actor's contribution to the risk and capacity to prevent it. Such an approach is more consistent with fairness than imposing responsibility exclusively according to the actor who happens to have the most direct relationship with the patient.

This model also has implications for evidentiary burdens. AI systems can create substantial information asymmetry. Patients may know that an incorrect diagnosis occurred but have no access to the algorithm's validation data, audit logs, version history, training limitations, or deployment documentation. Developers may possess the relevant technical evidence, while hospitals possess records concerning procurement, validation, training, and system monitoring. Physicians possess information about how the output influenced the clinical decision.

A meaningful liability regime must therefore preserve access to relevant evidence. Medical records should document material AI involvement in diagnosis, including the system used, the relevant output, the physician's clinical assessment, and significant deviations or overrides. Permenkes No. 24 of 2022 establishes the current Indonesian framework for medical records and emphasizes electronic medical records, security, and confidentiality in the context of digital healthcare. Although the regulation does not create a complete AI audit regime, its emphasis on electronic documentation provides a foundation upon which AI-related traceability requirements can be developed. Documentation is not merely administrative. It is essential to causation. Without a record of whether AI was used, how its output was interpreted, and whether the physician independently reviewed it, a patient may be unable to establish the causal pathway through which the injury occurred. AI governance therefore intersects directly with procedural justice.

Human Oversight as a Legal Duty Rather Than a Formality

A central component of the proposed framework is the recognition that human oversight should constitute a legal duty whenever AI materially influences clinical decisions. The concept of “human-in-the-loop” is insufficient if it means only that a physician remains formally present somewhere in the decision-making process. The relevant question is whether the physician has meaningful capacity to understand, question, and override the system.

The distinction is increasingly recognized in international governance. Article 14 of the EU AI Act requires high-risk AI systems to be designed so that natural persons can effectively oversee their operation. Human oversight must be proportionate to the system's risks and level of autonomy, and overseers must be capable of understanding limitations, recognizing excessive reliance, interpreting outputs, and deciding not to use or to override the system. The EU framework therefore treats human oversight as an operational capability rather than a symbolic presence.

The same principle should apply to Indonesian healthcare. A hospital should not satisfy its responsibility merely by assigning a physician to review AI-generated outputs if the physician has not been trained to understand the system's limitations or lacks authority to override the recommendation. Likewise, a physician should not be considered to have exercised independent judgment if institutional procedures effectively require automatic acceptance of algorithmic outputs.

This principle has direct consequences for malpractice. Where an AI system is used as a diagnostic aid, the physician's standard of care should include an obligation to evaluate whether the output is consistent with the patient's clinical presentation. This does not require the physician to reproduce the algorithm's analysis manually. It requires reasonable clinical scrutiny.

The obligation should become more demanding as the potential consequences of error increase. A low-risk AI tool used to prioritize routine administrative information does not require the same level of oversight as an algorithm recommending whether a patient should undergo an invasive procedure. The law should therefore apply a proportionality principle: greater clinical autonomy given to AI and greater potential harm from error should correspond to stronger human oversight.

The concept also provides a useful boundary for developer and institutional liability. If the system is designed in a way that prevents meaningful human intervention, the developer may bear responsibility for creating an unsafe architecture. If the hospital deploys the system without ensuring that physicians can adequately supervise it, institutional responsibility may arise. If the physician possesses the necessary capacity but fails to exercise it, professional negligence becomes more plausible. This approach avoids the false choice between human and machine responsibility. AI does not need to be treated as an independent legal actor for responsibility to be distributed. Legal responsibility can remain with natural persons and organizations while recognizing that the causal mechanism of harm is technologically mediated.

A Proposed Indonesian Model of AI-Assisted Medical Liability

The preceding analysis supports a layered model of liability built around the relationship between control, foreseeability, preventability, and causation. The physician should remain the primary professional decision-maker, but not the automatic bearer of every risk created by the technological ecosystem. The healthcare institution should assume responsibility for governance and deployment, while developers should bear responsibility for defects and failures within their technical sphere of control.

The first layer concerns clinical negligence. A physician should be liable where the clinician's selection, use, interpretation, or disregard of an AI system falls below the applicable professional standard and contributes causally to patient harm. Relevant circumstances may include failure to consider obvious clinical

contradictions, unjustified reliance on a system known to be inappropriate for the patient, failure to investigate an anomalous output, or failure to exercise independent professional judgment where the circumstances reasonably demanded it.

The second layer concerns technological defect. Responsibility should attach to the developer or provider where the AI system contains a design or performance defect, inadequate validation, material undisclosed limitation, misleading performance representation, or other technological deficiency that materially contributes to the injury. The focus should be on the provider's capacity to prevent the risk rather than on the mere fact that the algorithm produced an incorrect prediction.

The third layer concerns institutional governance failure. Healthcare facilities should be responsible where they deploy AI without adequate assessment of suitability, validation, monitoring, training, documentation, cybersecurity, or human oversight. This layer is particularly important because the hospital controls the environment in which the technology interacts with clinical practice. Article 303 of Law No. 17 of 2023, which assigns quality-control responsibility to healthcare facilities, provides an important statutory basis for this institutional dimension.

The fourth layer concerns shared causation. Where physician negligence, technological defect, and institutional failure interact, responsibility should not be artificially reduced to one category. Instead, the adjudicative process should identify the contribution of each actor to the injury and apply the relevant legal mechanisms to each. This is particularly important where the causal chain would not have produced the injury without the interaction of multiple failures.

Such a model does not require Indonesia to enact an entirely new AI-specific tort law immediately. Existing principles can accommodate a significant portion of the problem if interpreted in a technologically neutral but risk-sensitive manner. What is needed first is a clearer articulation of how those principles apply when clinical decisions are mediated by algorithms.

Regulation should nevertheless develop more explicit requirements for high-impact clinical AI. Healthcare institutions should be required to maintain documentation concerning the systems they deploy, their intended purposes, validation status, relevant limitations, responsible personnel, monitoring procedures, and incident-reporting mechanisms. Developers should provide sufficient technical and clinical documentation to enable institutions and physicians to use systems safely. Material AI-related incidents should be documented in a manner that permits later investigation.

The patient should also retain access to relevant information concerning AI involvement. This is not because every patient must understand the technical details of machine learning but because the patient's ability to seek redress depends upon knowing whether AI materially contributed to the decision. Transparency should therefore be understood as part of accountability.

Data governance is also inseparable from liability. AI-assisted diagnosis depends heavily on health data, which Indonesian Law No. 27 of 2022 concerning Personal Data Protection recognizes as a category of specific personal data. The legal protection of such information means that healthcare institutions must consider not only whether an AI system is clinically reliable but also whether patient data are processed lawfully and securely. The use of patient data for diagnosis, system monitoring, and secondary model development may involve distinct legal considerations.

The relationship between data protection and medical liability is particularly important when a data-quality problem contributes to diagnostic error. If an AI system produces systematically inaccurate results because the hospital's data are incomplete or improperly processed, the causal inquiry may extend beyond the algorithm itself. The relevant failure could lie in data governance, system design, clinical deployment, or several of these factors simultaneously. The regulatory objective should therefore be to create a traceable chain of responsibility. At each stage of the AI lifecycle, there should be an identifiable actor responsible for relevant safety decisions. This does not require assigning every possible risk to one party. It requires ensuring that no important risk exists within a legal vacuum.

From Individual Blame to Risk Governance

The deeper significance of AI-assisted medical liability is that it requires a shift from an exclusively individualistic conception of malpractice toward a governance-oriented conception of healthcare safety. Traditional malpractice litigation is often retrospective: a patient suffers harm, and the legal system asks whether an individual professional acted negligently. AI requires a complementary prospective question: whether the healthcare system was reasonably designed to prevent foreseeable technological failures.

This shift does not diminish the importance of individual accountability. On the contrary, physician responsibility remains essential because patients interact with healthcare professionals rather than algorithms as legal and ethical decision-makers. The point is that professional responsibility exists within an institutional and technological environment. If that environment systematically creates risks that physicians cannot reasonably control, assigning all responsibility to clinicians is neither fair nor effective. The World Health Organization's governance framework emphasizes that AI in health should protect human autonomy, promote safety, ensure transparency, establish accountability, and provide mechanisms for redress (World Health Organization 2021). These principles support a broader conception of responsibility in which AI governance is not merely about preventing technical malfunction but about creating institutional structures capable of identifying and correcting harmful outcomes.

The WMA's 2025 statement similarly emphasizes patient-centered, physician-led care while recognizing shared responsibilities among clinicians, healthcare

organizations, developers, regulators, and other stakeholders. It also stresses that mechanisms should exist for meaningful challenges to AI outputs and that human accountability must remain intact. This is particularly relevant to Indonesia because a layered liability model can be developed without granting AI independent legal personhood.

The idea of AI legal personhood is unnecessary for medical liability. A machine cannot meaningfully provide compensation, explain its conduct in the normative sense required by law, or bear institutional obligations. The relevant legal actors already exist: physicians, healthcare facilities, developers, manufacturers, insurers, and regulators. The challenge is to allocate responsibility among them according to their respective capacities and duties.

A governance-oriented model also changes the role of litigation. Litigation should remain available as a remedy, but the system should encourage earlier identification of AI-related incidents through internal reporting, clinical audit, professional review, and regulatory oversight. If an AI diagnostic system repeatedly produces errors, waiting until multiple patients are injured before intervention would be inconsistent with patient safety.

The requirement for incident reporting is therefore important. AI-related errors should be treated as opportunities for system correction rather than merely as evidence to be used in subsequent litigation. This requires institutions to develop reporting mechanisms that distinguish between ordinary clinical uncertainty and technology-related safety events. The purpose should be to identify recurring patterns of error and determine whether the problem lies in the model, data, user interaction, deployment environment, or institutional governance. Such an approach is particularly important because AI systems may fail systematically rather than randomly. A single physician error may affect one patient. A flawed algorithm may affect thousands of patients before the pattern becomes apparent. The potential scale of harm therefore justifies stronger institutional monitoring than would be required for an ordinary clinical tool.

The legal framework should consequently encourage organizations to preserve auditability. Records concerning model versions, updates, performance monitoring, significant incidents, and clinical overrides can become essential evidence in determining responsibility. Without such records, the complexity of AI may become a shield against accountability because every actor can plausibly deny knowledge of how the error occurred. The ultimate objective should be to ensure that technological complexity does not produce legal opacity. Patients should not bear the consequences of an accountability gap merely because the causal pathway involves software, data, and multiple organizations. Nor should physicians be transformed into guarantors of technologies they did not design and cannot reasonably inspect. The appropriate legal response is to reconstruct responsibility around the distribution of risk.

Conclusion

AI-assisted diagnosis exposes a structural limitation in conventional medical malpractice law: patient harm can no longer be assumed to originate exclusively from an identifiable human clinical decision. An incorrect diagnosis may result from physician negligence, algorithmic defect, biased or inadequate training data, inappropriate deployment, institutional governance failure, automation bias, or an interaction among several of these factors. Indonesian law already contains important foundations for addressing this problem. Law No. 17 of 2023 establishes patient safety obligations, professional-discipline mechanisms, dispute-resolution procedures, and institutional quality-control responsibilities, while Government Regulation No. 28 of 2024 develops the broader regulatory architecture for healthcare and health technology. The appropriate legal development is therefore not to abandon malpractice doctrine but to reconstruct it within a broader framework of distributed accountability. Physician liability should remain grounded in professional negligence; developer responsibility should address technological risks within the developer's sphere of control; healthcare institutions should bear responsibility for failures of selection, validation, deployment, training, monitoring, and governance; and cases involving multiple contributing failures should be analyzed through shared causation rather than forced into a single category of fault.

The central normative principle should be that liability follows the actor's capacity to prevent, identify, and mitigate the relevant risk. This principle avoids both extremes that threaten the legitimacy of AI in healthcare. Physicians should not become automatic guarantors of algorithmic accuracy simply because they remain responsible for clinical decisions, while developers should not receive immunity merely because their products operate through physicians. A sustainable Indonesian framework should instead require meaningful human oversight, institutional validation, technological transparency, incident documentation, and traceable allocation of responsibility across the AI lifecycle. The result would be a model in which innovation and accountability are not treated as opposing objectives. AI may improve diagnostic accuracy and expand clinical capacity, but its benefits cannot justify an accountability vacuum when patients are harmed. The legal task is therefore not to determine whether an AI-assisted diagnostic error is “malpractice” or “system failure” as mutually exclusive categories. It is to determine which combination of professional, technological, and institutional failures produced the injury and to assign responsibility to those actors best positioned to prevent the relevant risk. That layered approach would provide Indonesian patients with a more effective path to remedy while giving physicians, hospitals, and AI developers clearer incentives to build a safer digital healthcare system.

References

Amann, Julia, Alessandro Blasimme, Effy Vayena, and others. 2020. "Explainability for Artificial Intelligence in Healthcare: A Multidisciplinary Perspective." *BMC Medical Informatics and Decision Making* 20: 310. https://doi.org/10.1186/s12911-020-01332-6.

Beauchamp, Tom L., and James F. Childress. 2019. *Principles of Biomedical Ethics*. 8th ed. New York: Oxford University Press.

Burrell, Jenna. 2016. "How the Machine 'Thinks': Understanding Opacity in Machine Learning Algorithms." *Big Data & Society* 3 (1): 1–12. https://doi.org/10.1177/2053951715622512.

Cestonaro, Carlo, Alessandro Delicati, Beatrice Marcante, Luciana Caenazzo, and Pamela Tozzo. 2023. "Defining Medical Liability When Artificial Intelligence Is Applied on Diagnostic Algorithms: A Systematic Review." *Frontiers in Medicine* 10: 1305756. https://doi.org/10.3389/fmed.2023.1305756.

Char, Danton S., Nigam H. Shah, and David Magnus. 2018. "Implementing Machine Learning in Health Care—Addressing Ethical Challenges." *New England Journal of Medicine* 378 (11): 981–983. https://doi.org/10.1056/NEJMp1714229.

Cohen, I. Glenn, Andrew Slottje, and Sara Gerke. 2024. "Medical AI and Tort Liability." In *Artificial Intelligence in Medicine: From Ethical, Social, and Legal Perspectives*, 89–104. Academic Press. https://doi.org/10.1016/B978-0-323-95068-8.00007-8.

European Union. 2024. *Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act).* Official Journal of the European Union.

Faden, Ruth R., and Tom L. Beauchamp. 1986. *A History and Theory of Informed Consent*. New York: Oxford University Press.

Gianfrancesco, Milena A., Suzanne Tamang, Jino Yazdany, and Gabriela Schmajuk. 2018. "Potential Biases in Machine Learning Algorithms Using Electronic Health Record Data." *JAMA Internal Medicine* 178 (11): 1544–1547. https://doi.org/10.1001/jamainternmed.2018.3763.

Jobin, Anna, Marcello Ienca, and Effy Vayena. 2019. "The Global Landscape of AI Ethics Guidelines." *Nature Machine Intelligence* 1: 389–399. https://doi.org/10.1038/s42256-019-0088-2.

Kelly, Christopher J., Alan Karthikesalingam, Mustafa Suleyman, Greg Corrado, and Dominic King. 2019. "Key Challenges for Delivering Clinical Impact with Artificial Intelligence." *BMC Medicine* 17: 195. https://doi.org/10.1186/s12916-019-1426-2.

Kiseleva, Anastasia, and Paul De Hert. 2020. "If You Don't Know What You're Doing, Don't Do It! Artificial Intelligence and the Right to Explanation."

International Data Privacy Law 10 (2): 128–141. https://doi.org/10.1093/idpl/ipaa007.

Mello, Michelle M., and Neel Guha. 2024. "Understanding Liability Risk from Using Health Care Artificial Intelligence Tools." *New England Journal of Medicine* 390 (3): 271–278. https://doi.org/10.1056/NEJMhle2308901.

Morley, Jessica, Luciano Floridi, Libby Kinsey, and Anat Elhalal. 2020. "From What to How: An Initial Review of Publicly Available AI Ethics Tools, Methods and Research to Translate Principles into Practices." *Science and Engineering Ethics* 26: 2141–2168. https://doi.org/10.1007/s11948-019-00165-5.

Obermeyer, Ziad, Brian Powers, Christine Vogeli, and Sendhil Mullainathan. 2019. "Dissecting Racial Bias in an Algorithm Used to Manage the Health of Populations." *Science* 366 (6464): 447–453. https://doi.org/10.1126/science.aax2342.

Parikh, Ravi B., Michael J. Obermeyer, and others. 2021. "Artificial Intelligence and Liability in Medicine: Balancing Safety and Innovation." *Milbank Quarterly* 99 (3): 629–647. https://doi.org/10.1111/1468-0009.12504.

Price, W. Nicholson, Sara Gerke, and I. Glenn Cohen. 2021. "How Much Can Potential Jurors Tell Us About Liability for Medical Artificial Intelligence?" *Journal of Nuclear Medicine* 62 (1): 15–16. https://doi.org/10.2967/jnumed.120.257196.

Rajkomar, Alvin, Jeffrey Dean, and Isaac Kohane. 2019. "Machine Learning in Medicine." *New England Journal of Medicine* 380: 1347–1358. https://doi.org/10.1056/NEJMra1814259.

Republic of Indonesia. 2022a. *Law Number 27 of 2022 concerning Personal Data Protection*. Jakarta: Government of the Republic of Indonesia.

Republic of Indonesia. 2022b. *Regulation of the Minister of Health Number 24 of 2022 concerning Medical Records*. Jakarta: Ministry of Health of the Republic of Indonesia.

Republic of Indonesia. 2023. *Law Number 17 of 2023 concerning Health*. Jakarta: Government of the Republic of Indonesia.

Republic of Indonesia. 2024. *Government Regulation Number 28 of 2024 concerning the Implementing Regulation of Law Number 17 of 2023 concerning Health*. Jakarta: Government of the Republic of Indonesia.

Rudin, Cynthia. 2019. "Stop Explaining Black Box Machine Learning Models for High Stakes Decisions and Use Interpretable Models Instead." *Nature Machine Intelligence* 1: 206–215. https://doi.org/10.1038/s42256-019-0048-x.

Smith, Helen, and Kit Fotheringham. 2020. "Artificial Intelligence in Clinical Decision-Making: Rethinking Liability." *Medical Law Review* 28 (4): 636–657. https://doi.org/10.1093/medlaw/fwaa022.

Smith, Helen. 2021. "Clinical AI: Opacity, Accountability, Responsibility and Liability." *AI & Society* 36: 535–545. https://doi.org/10.1007/s00146-020-01019-6.

Tabassi, Elham. 2023. *Artificial Intelligence Risk Management Framework (AI RMF 1.0).* NIST AI 100-1. Gaithersburg, MD: National Institute of Standards and Technology. https://doi.org/10.6028/NIST.AI.100-1.

Tobia, Kevin, et al. 2021. "When Does Physician Use of AI Increase Liability?" *Journal of Nuclear Medicine* 62 (1): 17–21.

Topol, Eric. 2019. *Deep Medicine: How Artificial Intelligence Can Make Healthcare Human Again*. New York: Basic Books.

Vayena, Effy, Alessandro Blasimme, and I. Glenn Cohen. 2018. "Machine Learning in Medicine: Addressing Ethical Challenges." *PLoS Medicine* 15 (11): e1002689. https://doi.org/10.1371/journal.pmed.1002689.

World Health Organization. 2021. *Ethics and Governance of Artificial Intelligence for Health: WHO Guidance*. Geneva: World Health Organization.

World Medical Association. 2025. *WMA Statement on Artificial and Augmented Intelligence in Medical Care*. Ferney-Voltaire: World Medical Association.

***

Acknowledgment
None

Funding Information
None

Conflicting Interest Statement
The authors state that there is no conflict of interest in the publication of this article.

Publishing Ethical and Originality Statement
All authors declared that this work is original and has never been published in any form and in any media, nor is it under consideration for publication in any journal, and all sources cited in this work refer to the basic standards of scientific citation.

Generative AI Statement
N/A