Indonesian Health Justice Review
ISSN: XXXX-XXXXX (Online)
Vol. 3, No. 1, 2026, pp. 1-26
Published online since March 25, 2026

Who Is Liable When AI Misdiagnoses? Reconstructing the Legal Accountability of Physicians and AI Developers in Indonesian Healthcare

Abdul Rosyid*
Universitas Padjadjaran, Bandung, Indonesia
abd.rosyid@gmail.com

Fikri Nur Alamsyah*
Telkom University, Bandung, Indonesia
fikrinur-alam@gmail.com

DOI: https://doi.org/10.65815/qcrc7058

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

*Corresponding Author

Abstract

The increasing use of artificial intelligence (AI) in clinical decision-making presents significant challenges for healthcare liability law. AI-assisted diagnosis may improve efficiency and diagnostic accuracy, but errors generated by algorithmic systems raise a fundamental legal question: who should be held accountable when an AI-supported diagnosis causes patient harm? This article examines the allocation of legal responsibility between physicians, healthcare institutions, and AI developers when AI contributes to diagnostic errors in Indonesia. Using a normative juridical approach, the study analyzes Indonesian health law, medical liability principles, patient protection frameworks, and emerging approaches to AI governance. The analysis demonstrates that Indonesia's existing legal framework remains predominantly human-centered and does not clearly regulate liability arising from autonomous or semi-autonomous clinical technologies. Physicians may remain responsible for clinical decisions, yet imposing liability exclusively on physicians may be problematic where errors originate from defective algorithms, inadequate training data, or undisclosed system limitations. This article proposes a distributed liability framework based on the degree of human control, foreseeability of harm, system transparency, and the respective duties of physicians, healthcare providers, and AI developers. Such a framework is necessary to ensure patient protection without discouraging responsible technological innovation in healthcare.

[*Meningkatnya penggunaan kecerdasan buatan (AI) dalam pengambilan keputusan klinis menghadirkan tantangan signifikan bagi hukum pertanggungjawaban layanan kesehatan. Diagnosis yang dibantu AI dapat meningkatkan efisiensi dan akurasi diagnostik, tetapi kesalahan yang*

dihasilkan oleh sistem algoritmik menimbulkan pertanyaan hukum mendasar: siapa yang harus bertanggung jawab ketika diagnosis yang didukung AI menyebabkan kerugian pada pasien? Artikel ini mengkaji alokasi tanggung jawab hukum antara dokter, lembaga layanan kesehatan, dan pengembang AI ketika AI berkontribusi pada kesalahan diagnostik di Indonesia. Dengan menggunakan pendekatan yuridis normatif, studi ini menganalisis hukum kesehatan Indonesia, prinsip-prinsip pertanggungjawaban medis, kerangka perlindungan pasien, dan pendekatan baru terhadap tata kelola AI. Analisis menunjukkan bahwa kerangka hukum Indonesia yang ada sebagian besar masih berpusat pada manusia dan tidak secara jelas mengatur pertanggungjawaban yang timbul dari teknologi klinis otonom atau semi-otonom. Dokter mungkin tetap bertanggung jawab atas keputusan klinis, namun membebankan pertanggungjawaban secara eksklusif kepada dokter mungkin bermasalah jika kesalahan berasal dari algoritma yang cacat, data pelatihan yang tidak memadai, atau keterbatasan sistem yang tidak diungkapkan. Artikel ini mengusulkan kerangka pertanggungjawaban terdistribusi berdasarkan tingkat kendali manusia, kemampuan untuk memperkirakan kerugian, transparansi sistem, dan kewajiban masing-masing dokter, penyedia layanan kesehatan, dan pengembang AI. Kerangka kerja seperti itu diperlukan untuk memastikan perlindungan pasien tanpa menghambat inovasi teknologi yang bertanggung jawab dalam perawatan kesehatan.]

Keywords: Artificial intelligence, medical liability, misdiagnosis, patient protection, health law

Introduction

Artificial intelligence has increasingly become part of the technological infrastructure through which healthcare services are delivered, although its actual clinical integration remains considerably more complex than the rapid development of algorithmic capability might suggest. AI systems are now capable of assisting with image interpretation, risk prediction, disease classification, clinical prioritization, and other decision-support functions. Their potential attraction is particularly strong in healthcare systems facing shortages of specialists, uneven distribution of medical expertise, increasing diagnostic workloads, and demands for faster clinical decision-making. Yet technical performance in an experimental environment does not automatically translate into safe clinical effectiveness. AI systems can encounter dataset shifts, confounding variables, insufficient external validation, limited generalizability, and implementation problems when transferred from controlled research environments to heterogeneous clinical settings (Kelly et al. 2019). The legal significance of these limitations is considerable because a diagnostic error produced through AI cannot be understood simply as either a

machine error or a physician error. Rather, it may represent the outcome of an interconnected process involving data providers, software developers, healthcare institutions, clinicians, and patients. The emergence of AI therefore requires medical liability law to reconsider how responsibility should be allocated when technological and professional decisions jointly produce harm.

The conventional architecture of medical liability is principally organized around human conduct. A patient who suffers injury from an incorrect diagnosis traditionally asks whether the physician failed to satisfy the professional standard of care, whether that failure caused the injury, and whether legally recognizable damage resulted. AI complicates this structure because the physician may not independently generate every element of the diagnostic judgment. Instead, the physician may rely upon an algorithm that identifies a possible disease, assigns a probability, recommends further testing, or classifies medical images. Some machine-learning systems can produce highly accurate predictions without providing an intelligible account of how the prediction was generated. Price, Gerke, and Cohen observe that medical AI can therefore create difficult questions concerning the circumstances in which physicians should be liable when they rely on algorithmic recommendations, particularly where the physician cannot readily evaluate the internal reasoning of the system (Price, Gerke, and Cohen 2019). The legal difficulty is not simply technological opacity. It concerns whether the law can fairly attribute responsibility to a physician for a technological output whose underlying risks may have been determined before the physician ever encountered the patient. A physician-centered liability model may consequently produce accountability without genuine control.

The Indonesian context makes this problem especially significant because the country's contemporary health-law framework has undergone substantial consolidation. Law Number 17 of 2023 concerning Health replaced several earlier statutes, including Law Number 29 of 2004 concerning Medical Practice and Law Number 44 of 2009 concerning Hospitals, and now functions as the principal legislative framework for healthcare governance. The statute establishes professional obligations and legal protection for medical personnel and health workers while simultaneously recognizing institutional responsibility within hospitals. This architecture is important because it demonstrates that Indonesian health law already distributes some responsibility beyond the individual physician. Nevertheless, the legislation was not specifically drafted around machine-learning systems capable of generating diagnostic recommendations. Consequently, the application of existing principles to AI-assisted misdiagnosis requires interpretation rather than mechanical application. The central question becomes whether concepts such as professional negligence, hospital responsibility, patient protection, and technological safety can be interpreted sufficiently broadly to accommodate AI or whether Indonesia requires a distinct regulatory framework. This article argues that existing law provides a foundation, but not a complete

solution, and that judicial and legislative reconstruction is necessary to avoid both accountability gaps and disproportionate liability.

The problem is particularly acute because AI-assisted diagnosis involves several distinct forms of potential failure that may occur independently or simultaneously. An algorithm may be technically defective; a training dataset may inadequately represent the population; the developer may overstate performance; a hospital may deploy the system outside its validated purpose; a physician may misunderstand the output; or an institutional workflow may make meaningful human review practically impossible. Each of these circumstances creates a different legal problem. Treating them all as "medical malpractice" obscures important distinctions between professional negligence, product defect, institutional negligence, data governance failure, and implementation error. Conversely, characterizing every incorrect prediction as a technological defect would be equally problematic because statistical systems necessarily produce false positives and false negatives even when properly designed and validated. The appropriate legal inquiry must therefore move beyond the binary question of whether the AI was correct or incorrect. It should instead determine which actor created the relevant risk, which actor had the ability to identify or mitigate that risk, whether the risk was foreseeable, and whether a legally relevant duty was breached. This approach is consistent with the broader proposition that accountability for healthcare AI should be distributed among stakeholders rather than assigned exclusively to the professional who happens to deliver the final clinical decision (Gerke, Minssen, and Cohen 2020).

The concept of distributed responsibility is also supported by international governance approaches. The World Health Organization has emphasized that AI in health should preserve human autonomy, promote safety and public benefit, ensure transparency and intelligibility, foster responsibility and accountability, promote inclusiveness and equity, and remain responsive and sustainable (World Health Organization 2021). These principles have direct implications for liability because they identify responsibility as a characteristic of the entire AI ecosystem rather than merely of the healthcare worker who interacts with the final output. The WHO further recognizes that systems trained predominantly on data from high-income countries may not perform adequately in different socioeconomic and healthcare settings. For Indonesia, this observation is particularly relevant because clinical AI imported from foreign markets may be developed and validated using populations, disease patterns, equipment, workflows, or health-system conditions that differ from those encountered domestically. The resulting legal question is therefore not simply whether the developer's algorithm achieved a particular accuracy rate. It is whether the developer and deploying institution exercised reasonable care in determining whether the system was fit for the Indonesian clinical environment in which patients would actually be exposed to its risks.

The legal scholarship on AI-assisted healthcare has identified liability as one of the principal unresolved issues accompanying technological adoption. Gerke,

Minssen, and Cohen identify safety and effectiveness, liability, privacy, cybersecurity, and intellectual property as interconnected legal challenges rather than isolated regulatory categories (Gerke, Minssen, and Cohen 2020). This interconnectedness matters because an AI misdiagnosis may originate from a combination of factors. For example, a developer may train a model on data collected without sufficient attention to population diversity; a hospital may integrate the model into an electronic medical-record system without adequate validation; and a physician may then interpret its output as more authoritative than warranted. A patient's resulting injury cannot easily be reduced to one discrete act of negligence. Legal analysis should therefore reconstruct the chain of responsibility rather than search automatically for a single wrongdoer. The proposed approach in this article is not intended to dilute professional accountability. Instead, it seeks to align legal responsibility with actual capacity to prevent harm. An actor who creates or controls a particular technological risk should generally bear responsibility for that risk when the relevant legal elements of liability are established.

The research gap addressed by this article lies in the intersection between Indonesia's developing health-law framework and the emerging problem of algorithmic clinical responsibility. Existing discussions of medical AI in Indonesia have increasingly examined ethics, data protection, digital health, and general regulatory uncertainty, but the specific allocation of liability among physicians, hospitals, and AI developers remains insufficiently reconstructed through an integrated legal framework. The difficulty is heightened by the fact that Indonesia's principal health legislation does not contain a dedicated chapter establishing liability for AI-generated medical decisions. The absence of an explicit provision, however, does not necessarily mean that existing law is incapable of responding. Indonesian legal interpretation can employ existing principles of professional responsibility, institutional liability, consumer and patient protection, tort principles, contractual obligations, and data governance. The more important question is how these principles should be coordinated. This article therefore asks three questions: first, how should Indonesian law characterize AI-assisted misdiagnosis; second, when should physicians, healthcare institutions, and AI developers bear responsibility; and third, what legal reconstruction can distribute liability fairly while maintaining patient protection and technological innovation?

Methodologically, this article employs a normative juridical approach using statutory, conceptual, and comparative approaches. Primary legal materials consist principally of Law Number 17 of 2023 concerning Health, Law Number 27 of 2022 concerning Personal Data Protection, Regulation of the Minister of Health Number 24 of 2022 concerning Medical Records, and Government Regulation Number 28 of 2024 concerning the implementation of the Health Law. Law Number 17 of 2023 is particularly significant because it currently provides the principal statutory foundation for medical and hospital responsibility in Indonesia, while the 2022 Medical Records Regulation provides the regulatory context for electronically documented clinical information. Secondary materials include peer-reviewed

scholarship concerning AI safety, explainability, algorithmic bias, physician liability, and healthcare governance. Comparative analysis is used not to transplant foreign law into Indonesia but to identify regulatory principles that may illuminate gaps in the Indonesian framework. The article ultimately proposes a distributed liability model organized around human control, risk creation, foreseeability, transparency, validation, and causal contribution.

The central thesis is that Indonesia should not construct AI medical liability around the assumption that the physician is always the legally responsible actor merely because a physician makes the final clinical decision. Such an approach would overlook the fact that AI-related risks may be embedded in software design, training data, validation procedures, procurement decisions, system configuration, and institutional workflow. At the same time, developers should not be transformed into universal insurers against every adverse clinical outcome merely because their technology was involved. The appropriate solution is a differentiated liability framework in which each actor is assessed according to the risk that actor created or controlled and the duty that actor owed to the patient or healthcare system. Physicians should remain accountable for unreasonable clinical conduct; hospitals should bear responsibility for failures of technology governance; and developers should bear responsibility for defects and failures within their technical and commercial sphere of control. Where these forms of responsibility overlap, concurrent liability should be available. Such a model can strengthen patient protection without creating a legal environment in which technological uncertainty automatically becomes physician fault.

AI-Assisted Diagnosis and the Transformation of Medical Negligence

The first analytical difficulty is distinguishing an incorrect AI output from legally actionable medical negligence. An AI system is fundamentally probabilistic and may produce an incorrect result even when operating exactly as designed. Diagnostic medicine itself also operates under uncertainty, and physicians can make reasonable clinical judgments that subsequently prove incorrect. Accordingly, neither an adverse outcome nor an AI error should automatically establish negligence. Medical negligence requires an assessment of the applicable duty and whether the relevant actor departed from that duty under the circumstances. This distinction becomes more important when an AI system is involved because the clinical outcome may result from an interaction between algorithmic performance and human judgment. Kelly et al. emphasize that clinical AI faces problems involving generalizability, dataset shift, bias, and implementation that may not become apparent from technical performance metrics alone (Kelly et al. 2019). The legal consequence is that the standard of care must consider the actual conditions under which the technology was used. A physician who reasonably uses a validated system within its intended purpose should not automatically be negligent when the system produces an unforeseeable false

negative. Conversely, a physician who ignores obvious contradictory clinical evidence cannot necessarily avoid responsibility simply by stating that the AI produced the erroneous recommendation.

Indonesian law provides a useful starting point for this distinction through the professional duties and protections established under Law Number 17 of 2023. The statute's structure indicates that professional responsibility is connected to compliance with professional standards, professional service standards, standard operating procedures, professional ethics, and patient needs, while legal protection is available to medical personnel who perform their duties consistently with those requirements. This orientation is fundamentally conduct-based rather than outcome-based. In the AI context, therefore, the question should be whether the physician acted reasonably in understanding, evaluating, and applying the AI output. A physician should not be expected to guarantee the mathematical correctness of a model that the physician did not design and cannot independently audit. However, neither should the use of AI eliminate the physician's professional duty to assess the patient as a whole. The appropriate standard is one of reasonable AI-assisted clinical judgment. That standard requires the physician to understand the system's intended function, recognize material limitations, evaluate whether the patient's presentation falls within the system's validated context, and exercise independent judgment where clinical circumstances contradict the algorithmic output. Such a standard preserves human responsibility without making physicians guarantors of algorithmic performance.

The emergence of AI nevertheless changes what constitutes reasonable professional conduct. A physician's duty cannot remain entirely static when technology changes the information available at the point of care. If a particular AI diagnostic system becomes widely accepted, independently validated, and demonstrably more effective than conventional diagnostic approaches for a specific condition, the professional standard may eventually evolve to expect physicians to use that technology in appropriate cases. Failure to use an established diagnostic aid could potentially become relevant to negligence analysis in the same way that failure to use a recognized diagnostic test may become professionally significant. Conversely, physicians should not be required to use experimental or poorly validated AI merely because it is technologically sophisticated. Price, Gerke, and Cohen highlight the difficulty of determining when physicians should rely on AI and how the standard of care may respond to increasingly capable decision-support systems (Price, Gerke, and Cohen 2019). The appropriate legal test should therefore remain technologically neutral while being sensitive to the actual state of medical knowledge. The question should not be whether the physician used AI, but whether the physician made a reasonable clinical decision considering all reliable information reasonably available at the relevant time. This formulation prevents technological adoption from becoming either a shield or an automatic basis for liability.

A particularly important issue is automation bias, meaning the tendency of human decision-makers to place excessive confidence in automated recommendations. In clinical practice, the risk can become significant when an AI system is marketed as highly accurate or integrated into a workflow in a way that visually or procedurally prioritizes its recommendations. The physician may technically retain the authority to override the AI but practically experience institutional pressure to follow it. In such circumstances, assigning full responsibility to the physician merely because the final click or signature belongs to the physician would ignore the architecture of the decision environment. The legal assessment should instead consider whether the system was presented as authoritative, whether appropriate warnings were supplied, whether contradictory clinical information was visible, and whether the physician had sufficient time and resources for independent assessment. This does not excuse professional negligence. It recognizes, however, that human oversight is meaningful only if the human operator possesses sufficient information, competence, authority, and opportunity to exercise judgment. WHO's emphasis on maintaining human control over medical decisions therefore cannot be interpreted as simply placing a human at the end of an automated pipeline; meaningful human autonomy requires institutional conditions that make independent judgment realistically possible (World Health Organization 2021).

The problem of algorithmic bias further demonstrates why the physician cannot always be regarded as the primary creator of diagnostic risk. Obermeyer et al. demonstrated that a widely used healthcare algorithm produced substantial racial disparities because it predicted healthcare expenditure rather than illness, thereby embedding historical inequalities in access to healthcare into the algorithmic outcome (Obermeyer et al. 2019). The importance of this example extends beyond racial classification. It illustrates that algorithmic outputs can reflect design choices that are invisible to clinicians using the final system. A physician may observe the output without knowing what proxy variable was selected during model development, how the training population was constructed, or which cases were excluded. If an algorithm systematically underdiagnoses a particular population because its training data inadequately represent that population, it would be conceptually incorrect to treat every resulting clinical error as an independent physician failure. The developer's choice of dataset and outcome variable may have created the underlying risk. The hospital may additionally contribute if it fails to conduct local validation. Physician responsibility should therefore depend upon whether the relevant bias or limitation was reasonably detectable by a clinician at the point of care. Where it was not, developer and institutional responsibility becomes substantially more important.

Clinical AI also raises the question of what constitutes adequate validation. Technical developers may report sensitivity, specificity, accuracy, area under the receiver operating characteristic curve, or other statistical measures. Yet these measures do not necessarily establish that an AI system is safe for every clinical

population or workflow. Kelly et al. emphasize the importance of independent, local, and representative testing and warn against dataset shift and limited generalizability when AI moves into new clinical settings (Kelly et al. 2019). principle has direct relevance to Indonesia. A system developed using data from hospitals in North America or Europe may achieve impressive performance in its original validation population but behave differently when confronted with Indonesian patients, equipment, documentation practices, or disease distributions. Consequently, the mere existence of foreign regulatory approval or published accuracy data should not necessarily establish fitness for every Indonesian deployment. The hospital that introduces such technology should conduct appropriate due diligence concerning intended use and clinical context. The developer, meanwhile, should disclose the population and conditions in which the system has actually been validated. Liability should therefore follow the allocation of knowledge and control: developers are better positioned to know the technical limits of the model, while hospitals are better positioned to know the characteristics of their own clinical environment.

The concept of foreseeability provides an important doctrinal bridge between technical uncertainty and legal responsibility. Not every failure of a complex algorithm is foreseeable, and requiring developers to anticipate every possible model failure would impose an unrealistic standard. Nevertheless, foreseeable risks should not be externalized to patients merely because machine learning is technically complicated. If a developer knows that a model performs poorly on a relevant patient subgroup, has received reports of recurrent false negatives, or has not been validated for a clinical context in which the product is being marketed, the developer's failure to disclose or mitigate the problem can become legally significant. The same principle applies to hospitals. If a hospital receives warnings that a system has poor performance in a particular population but continues deploying it without safeguards, institutional responsibility may arise. Foreseeability should therefore be assessed not merely according to what the individual physician knew, but according to what each actor reasonably knew or should have known within that actor's sphere of expertise and control. This approach prevents technical complexity from becoming an accountability vacuum while avoiding the opposite extreme of imposing absolute liability for genuinely unforeseeable technological failures.

Physician and Hospital Liability under Indonesian Health Law

Law Number 17 of 2023 provides the primary statutory foundation for examining physician and hospital responsibility in Indonesia. Its significance lies not merely in consolidating earlier health legislation but in establishing an integrated institutional structure in which professional responsibility and healthcare-facility responsibility coexist. The statute expressly repealed Law Number 29 of 2004 concerning Medical Practice and Law Number 44 of 2009 concerning Hospitals,

among other health statutes. This legislative consolidation creates an opportunity to interpret professional liability in light of contemporary healthcare technologies rather than treating the older physician-centered model as fixed. Yet the absence of express provisions addressing AI means that the statute must be applied through its general concepts. The key interpretive question is whether the duties imposed on physicians and hospitals can encompass reasonable technological governance. This article argues that they can. Professional standards should be understood dynamically, while hospital responsibility should extend to organizational decisions that materially affect patient safety. Such interpretation is preferable to constructing a new legal category of "AI liability" detached from existing health-law principles because it maintains doctrinal continuity while adapting those principles to the technological environment.

The physician's position remains central because AI does not ordinarily replace the legal identity of the medical professional responsible for patient care. The physician is the actor who examines the patient, obtains clinical information, interprets diagnostic results, communicates with the patient, and decides upon treatment. Consequently, a physician cannot simply transfer professional responsibility to a software system by stating that "the AI made the diagnosis." Such an approach would be inconsistent with the professional nature of medical practice. Where the AI output is clearly inconsistent with clinical evidence and the physician nevertheless adopts it without reasonable justification, professional negligence may arise. Similarly, a physician who knowingly uses an AI system outside its intended clinical purpose or ignores explicit warnings about its limitations may have breached a professional duty. The legal significance of AI is therefore not to eliminate physician accountability but to redefine the content of reasonable clinical conduct. Physicians must become competent users of relevant technologies while retaining independent judgment. The appropriate standard is neither blind trust nor categorical distrust of AI. Instead, it is informed reliance, meaning reliance proportionate to the system's demonstrated reliability, clinical purpose, available evidence, and the physician's ability to identify circumstances in which independent assessment is required.

At the same time, the principle of physician responsibility should not be transformed into a doctrine of automatic liability. This distinction is essential for fairness and innovation. Suppose a physician uses an AI system that the hospital has approved, the developer has adequately validated, and the relevant regulatory authorities have accepted for the intended clinical purpose. The physician examines the patient, considers the medical history, reviews the AI output, and acts consistently with prevailing professional practice. The system nevertheless produces a rare and unforeseeable false negative. Treating the physician as negligent merely because the patient was harmed would transform professional liability into strict outcome liability. Such a standard would be inconsistent with the general logic of professional negligence and could produce defensive medicine. Physicians might avoid AI systems even when they improve average diagnostic

performance because the legal consequences of an isolated error would be too uncertain. The better approach is to distinguish a bad outcome from a breach of duty. Where the physician's conduct was reasonable and the relevant technological failure was outside the physician's reasonable capacity to detect or prevent, responsibility should shift toward the actors who controlled the underlying technological risk.

Hospital liability is more complex because hospitals occupy both clinical and organizational positions. A hospital does not merely provide a physical location for physicians. It selects technologies, negotiates contracts, establishes clinical protocols, controls information systems, provides staff training, monitors quality, and determines how new technologies are integrated into patient care. These activities create independent institutional risks. If a hospital purchases an AI diagnostic system without examining its validation evidence, permits use beyond the system's intended population, fails to train clinicians, or does not establish procedures for responding to algorithmic errors, the resulting harm cannot fairly be attributed solely to the physician who happens to interact with the software. The hospital has made organizational decisions that shaped the conditions under which the clinical error occurred. This supports interpreting hospital responsibility under Indonesian health law in a manner that encompasses technological governance rather than limiting it to traditional supervision of human employees.

Article 193 of Law Number 17 of 2023 is particularly significant in this regard because it establishes legal responsibility for hospitals for losses resulting from negligence committed by health human resources within the hospital. The provision provides an important statutory basis for institutional responsibility even though it was not expressly designed for AI. The analytical challenge is determining whether the concept should be limited to situations in which a particular health worker commits negligence or whether it can support a broader organizational interpretation. This article favors the latter, while recognizing that the precise doctrinal boundaries remain subject to judicial interpretation. In an AI environment, hospital negligence may occur through procurement, deployment, supervision, or failure to monitor the technology. If the hospital creates a workflow in which physicians are expected to follow an AI system without adequate opportunities for independent review, the institution has contributed to the risk. Likewise, if the hospital continues to deploy a system after evidence indicates substantial performance problems, institutional responsibility may arise independently of any particular physician's conduct. Article 193 should therefore be understood as a foundation for recognizing the hospital as an active participant in the allocation of technological risk.

This institutional interpretation is strengthened by the fact that healthcare AI cannot safely operate without governance structures. WHO's framework identifies accountability as extending to stakeholders responsible for ensuring that AI is used under appropriate conditions and by appropriately trained persons (World Health Organization 2021). The implication is that human oversight must be

operationalized through institutional systems. A hospital cannot satisfy its governance duty merely by telling physicians that they remain responsible for all decisions. It must provide training, define permissible uses, establish escalation procedures, maintain system logs, monitor performance, and ensure that physicians can challenge algorithmic recommendations. Otherwise, the formal assignment of responsibility to the physician may conceal a structural absence of meaningful oversight. The law should therefore examine whether the institution created the conditions necessary for responsible AI use. This approach is consistent with modern patient-safety thinking because many healthcare injuries arise from systemic conditions rather than isolated individual mistakes. AI makes that systemic dimension more visible because the technology itself is embedded within organizational decisions concerning procurement, configuration, integration, and supervision.

Electronic medical records are particularly important to this institutional accountability model. Regulation of the Minister of Health Number 24 of 2022 establishes the regulatory framework for medical records in Indonesia and remains in force. When AI contributes to a diagnosis, the medical record should ideally permit reconstruction of the relevant decision process. A conventional record may state the physician's final diagnosis and treatment, but an AI-assisted record may need additional information concerning the system used, the relevant output, material warnings, and whether the physician accepted or rejected the recommendation. Without such information, subsequent litigation may be unable to determine whether the error originated from the physician, the AI system, or the hospital's implementation of the technology. This is not merely an evidentiary inconvenience. It creates a structural disadvantage for patients because the relevant technological information is often controlled by institutions or vendors. Therefore, AI governance should require sufficient auditability to permit reconstruction of clinically significant decisions. The purpose is not to expose proprietary source code but to preserve evidence necessary for patient safety, quality improvement, regulatory oversight, and fair adjudication.

Hospital responsibility should also encompass post-deployment monitoring. A hospital cannot reasonably treat AI procurement as a one-time transaction because performance may vary after deployment. Changes in patient populations, clinical workflows, software versions, equipment, data quality, or surrounding systems may affect outcomes. Kelly et al. emphasize that regulation must be accompanied by thoughtful post-market surveillance because clinical safety cannot be determined solely at the moment of initial deployment (Kelly et al. 2019). The legal implication is that institutional duty should continue throughout the lifecycle of the system. A hospital that discovers a pattern of false negatives but does nothing may have a stronger basis for liability than a hospital confronted with a genuinely unforeseeable isolated error. Similarly, a developer that releases an update without adequate safety assessment may remain responsible for resulting defects. Lifecycle responsibility therefore provides a bridge between medical negligence and

technology governance, ensuring that accountability does not end once a software product enters a hospital.

AI Developer Liability: Defect, Data, Transparency, and Control

Developer responsibility represents the most significant departure from traditional medical malpractice because software developers generally have no direct physician-patient relationship. Nevertheless, the absence of a therapeutic relationship does not mean that developers are legally irrelevant. Developers design the algorithm, select or influence training data, establish performance objectives, determine intended uses, conduct validation, create documentation, and control software updates. These activities may directly determine the risk profile of the technology. If an AI system produces systematic diagnostic errors because of defective design or inadequate validation, it would be difficult to justify a legal framework under which only the physician bears responsibility. The developer is often the actor with the greatest technical capacity to identify and mitigate the relevant defect. The WHO has specifically emphasized the need for developers and other stakeholders to ensure safety, accuracy, transparency, and accountability throughout the AI lifecycle (World Health Organization 2021). Developer liability should therefore be based on the relationship between the developer's control over technological risks and the harm those risks cause, rather than on the existence of a direct doctor-patient contract.

The first category is defective design. An AI diagnostic system may be defective if its architecture predictably generates unsafe outputs for the clinical purpose for which it was marketed. However, the legal concept of defect must be carefully constructed because AI systems are probabilistic. A false negative does not automatically mean that the software is defective. The relevant question is whether the system performed within the safety and performance characteristics reasonably expected for its intended use. If a developer markets a system as appropriate for detecting a disease but testing reveals materially inadequate sensitivity in the relevant patient population, failure to disclose or correct the problem may support liability. Conversely, if the developer clearly identifies a known limitation and the system is used within the disclosed conditions, responsibility may shift toward the deploying institution or physician. This distinction is essential to prevent AI product liability from becoming absolute liability. The legal system should recognize that technological performance has statistical limits while simultaneously preventing developers from externalizing known or foreseeable risks onto patients.

Training data constitute a second potential source of developer responsibility. Machine-learning systems learn patterns from data, and the characteristics of those data can influence subsequent outputs. If a model is trained on a population that inadequately represents the population in which it will be deployed, its performance may deteriorate. Bias can also arise when the target variable is an

inappropriate proxy for the medical condition being predicted. Obermeyer et al. provide a particularly important example of how a seemingly rational predictive target—healthcare expenditure—can reproduce structural disparities because expenditure does not necessarily correspond to illness burden (Obermeyer et al. 2019). The legal lesson is that developers should not be judged solely by headline accuracy metrics. The underlying choice of data, outcome variables, validation population, and performance thresholds may be relevant to whether the system was reasonably designed. Where a developer could reasonably identify a material limitation and failed to disclose or mitigate it, the developer's responsibility becomes stronger.

The Indonesian context intensifies this concern because healthcare data may reflect distinctive demographic, epidemiological, institutional, and geographic conditions. A model trained elsewhere may encounter different disease prevalence, diagnostic pathways, languages, documentation practices, or medical equipment when deployed in Indonesia. The WHO expressly warns that AI systems trained primarily using data from high-income countries may not perform adequately in low- and middle-income settings (World Health Organization 2021). This does not establish that foreign-developed AI is inherently unsuitable for Indonesia. It establishes a reason for validation. A developer seeking to market an AI system for Indonesian clinical use should disclose the population on which the model was trained and validated and identify material limitations concerning generalizability. Hospitals, in turn, should not assume that foreign validation automatically establishes local fitness. The resulting responsibility is therefore shared but differentiated: developers are responsible for transparent representation of technical evidence, while hospitals are responsible for determining whether that evidence is sufficient for their actual clinical environment.

Transparency is another essential component of developer accountability. Clinical users cannot exercise meaningful oversight if they do not know what the AI system is designed to do, how well it performs, where it has limitations, or under what circumstances its output should not be relied upon. WHO identifies transparency, explainability, and intelligibility as important governance principles, while also emphasizing accountability and mechanisms for redress (World Health Organization 2021). However, transparency should not be equated with mandatory disclosure of source code. A developer may legitimately have proprietary interests in algorithms, model weights, and technical architecture. The legal requirement should instead focus on information necessary for safe clinical use and accountability. This may include intended use, validation population, clinically relevant performance measures, known failure modes, limitations, warnings, update history, and procedures for incident reporting. Such functional transparency would allow physicians and hospitals to make informed decisions without necessarily destroying intellectual property protection. Where a developer withholds material information that would have affected a reasonable hospital's decision to deploy the system, that omission may become relevant to liability.

Explainability must similarly be approached with caution. The assumption that every AI decision can or should be translated into a simple human-readable explanation is technically and conceptually problematic. Ghassemi, Oakden-Rayner, and Beam argue that current explainability techniques may not reliably provide the kind of patient-level insight necessary to justify clinical trust and caution against treating explainability as a substitute for rigorous validation (Ghassemi, Oakden-Rayner, and Beam 2021). Markus, Kors, and Rijnbeek likewise conclude that explainability can contribute to trustworthy AI but should be accompanied by measures such as data-quality reporting, extensive external validation, and regulation (Markus, Kors, and Rijnbeek 2021). Accordingly, Indonesian law should not make developer liability depend on whether a system possesses a particular form of explainable architecture. Instead, it should require sufficient information to permit safe use, meaningful oversight, and post-incident investigation. The central legal concern is not whether a physician can reproduce the model's mathematics but whether relevant actors can determine the system's intended purpose, reliability, limitations, and appropriate boundaries.

Developer responsibility should also extend to software updates where the developer retains control over the technology after deployment. Unlike conventional physical products, software can change after it has entered clinical use. An update may modify model behavior, introduce new functionality, alter performance, or address one risk while creating another. The European Union's Directive (EU) 2024/2853 provides an instructive comparative example by recognizing that manufacturers may remain responsible for defects arising after a product has entered the market when software, updates, upgrades, or machine-learning algorithms remain within their control. Although Indonesia should not automatically adopt the European model, the underlying principle is persuasive: legal responsibility should follow continuing control over technological risk. If a developer retains the ability to modify an AI system remotely, responsibility should not necessarily end at the original sale or deployment date. This is particularly important for cloud-based AI and software-as-a-service models in which the system is continuously maintained by the developer. Lifecycle control therefore provides a principled basis for extending developer duties beyond initial product delivery.

Cybersecurity creates another dimension of developer responsibility. A diagnostic AI system may become unsafe not because its underlying model is inaccurate but because unauthorized actors manipulate data, software, interfaces, or connected systems. Developers who design systems without reasonable security safeguards may contribute to patient harm. Healthcare institutions likewise have responsibilities because they control network infrastructure, access privileges, and operational security. The appropriate liability analysis should therefore identify which actor controlled the vulnerability and whether the vulnerability was reasonably foreseeable. This again supports the broader distributed-liability model. A technological product should not be treated as a static object for purposes of legal responsibility when its safety depends on continuing interaction between software,

networks, data, users, and institutional infrastructure. Gerke, Minssen, and Cohen identify cybersecurity as one of the central legal challenges of AI-driven healthcare, demonstrating that technological liability cannot be isolated from broader information-governance obligations (Gerke, Minssen, and Cohen 2020).

Developer liability should nevertheless have clear boundaries. A developer should not be held responsible simply because a physician intentionally uses an AI system outside its intended purpose, disregards explicit warnings, or materially alters the system without authorization. Nor should a developer be responsible for every adverse medical outcome where the product performed within its validated range and the injury resulted from independent clinical negligence. The appropriate test is therefore not whether the AI was involved in the chain of events but whether a defect or breach attributable to the developer materially contributed to the harm. This distinction preserves incentives for innovation. If developers face unlimited liability for every false prediction, they may have strong incentives to withdraw clinically beneficial systems or avoid high-risk diagnostic applications. Conversely, if liability is too weak, developers may have insufficient incentives to invest in validation, data quality, safety engineering, transparency, and post-market monitoring. The law should therefore impose responsibility proportionate to technical control and causal contribution rather than creating either blanket immunity or absolute liability.

Toward a Distributed Liability Framework for AI Misdiagnosis in Indonesia

The preceding analysis demonstrates that AI-assisted medical liability cannot adequately be reconstructed through a single-actor model. The physician, hospital, developer, data controller, and technology integrator may each occupy different positions in the causal chain. Their duties are also different because their capacities are different. A physician controls the immediate clinical judgment but generally cannot inspect the model's training data. A hospital controls procurement and implementation but may lack the technical capacity to redesign the algorithm. A developer controls model design and software updates but generally cannot control how an individual physician evaluates a patient. Liability should therefore reflect these differences. This article proposes a distributed liability framework based on six factors: human control, risk creation, foreseeability, duty, transparency, and causal contribution. The model can be expressed conceptually as follows: liability increases where an actor had greater control over the relevant risk, could reasonably foresee the harm, breached a specific duty, and materially contributed to the injury. Conversely, liability should decrease where the actor complied with applicable standards, lacked reasonable control over the relevant risk, and could not reasonably foresee or prevent the harm.

The first factor, human control, addresses the question of whether the actor had a realistic ability to prevent or mitigate the relevant error. Human control

should not be measured merely by asking who made the final decision. A physician may formally control the clinical decision but lack meaningful control over the algorithm's design. A hospital may control deployment but lack control over proprietary software architecture. A developer may control the model but lack control over whether a physician follows the output. The relevant question is therefore functional control. This approach prevents the common mistake of equating formal decision authority with actual risk control. WHO's principle of human autonomy is useful here because it emphasizes that humans should remain in control of healthcare systems and medical decisions, but meaningful control requires adequate institutional and technological conditions (World Health Organization 2021). A physician should consequently be judged on what the physician could reasonably control, while the developer and hospital should be judged on the risks they were respectively capable of controlling.

The second factor, risk creation, identifies which actor introduced or materially increased the risk that ultimately resulted in harm. Developers generally create risks through algorithm design, data selection, validation methodology, software architecture, and representations concerning system performance. Hospitals create risks through procurement, configuration, implementation, staffing, training, and monitoring. Physicians create risks through clinical misuse, unreasonable reliance, failure to review, or disregard of contradictory information. This factor provides a principled basis for differentiating concurrent responsibility. If a developer creates an unsafe model but a hospital deploys it without validation and a physician subsequently ignores clear clinical evidence, each actor may have contributed to the final injury. The law should not require an artificial choice among them. Instead, the causal inquiry should identify each materially contributing breach. Risk creation also has a preventive function: actors will have incentives to improve safety if they know that responsibility follows risks within their sphere of control. This is preferable to a model in which every participant assumes that another actor will ultimately bear responsibility.

The third factor is foreseeability. Foreseeability is particularly important because AI systems are probabilistic and technologically complex. A developer cannot reasonably predict every possible failure of a sophisticated model, just as a physician cannot identify every latent software defect. Legal responsibility should therefore concentrate on risks that were reasonably foreseeable based on available knowledge, validation evidence, warnings, prior incidents, or obvious limitations. Kelly et al. emphasize that developers must account for dataset shift, unintended bias, generalizability, and real-world implementation when translating AI systems into clinical practice (Kelly et al. 2019). These are precisely the categories of risk that may become foreseeable through appropriate testing. If a developer fails to conduct reasonable testing and consequently remains unaware of a foreseeable limitation, the absence of actual knowledge should not automatically eliminate responsibility. Conversely, where a defect could not reasonably have been identified using the scientific and technical knowledge available at the relevant

time, imposing liability may be unfair. Foreseeability therefore creates a balance between patient protection and innovation.

The fourth factor is transparency. In conventional medical negligence, patients can often obtain information about the physician's conduct through medical records. AI creates additional information asymmetries because critical evidence may be located within software systems, vendor documentation, training records, or proprietary logs. A patient may know only that an AI-assisted diagnosis was wrong while having no practical means to determine why it was wrong. This imbalance can make formal legal rights ineffective. Transparency should therefore operate as an evidentiary and governance obligation. Developers should preserve relevant information about system performance, intended use, limitations, and material updates. Hospitals should maintain records concerning deployment, training, configuration, and clinically significant AI outputs. Physicians should document material reliance upon or departure from algorithmic recommendations. Markus, Kors, and Rijnbeek emphasize that trustworthy healthcare AI requires not merely explainability but also data-quality reporting, external validation, and regulation (Markus, Kors, and Rijnbeek 2021). Transparency should therefore be understood broadly as the availability of information necessary to evaluate responsibility, rather than as an unrealistic requirement of complete technical openness.

The fifth factor is causal contribution. AI-related medical harm may involve multiple causal links, and conventional "but-for" reasoning may be insufficient when several actors contribute to the same injury. Suppose a developer releases a model with inadequate validation, a hospital deploys it without local testing, and a physician follows the output despite ambiguous symptoms. The injury may not have occurred in precisely the same way if any one of these conditions had been absent. The legal system should therefore identify substantial causal contributions rather than insist upon a single exclusive cause. This does not mean that every participant in the chain becomes liable. The claimant should still establish legally relevant causation between the defendant's breach and the harm. But where several breaches materially contribute, concurrent liability should be possible. Such an approach is especially important for AI because the technology often acts as a mediating mechanism between institutional and clinical decisions. The algorithm is not necessarily an independent cause; it may be the channel through which several earlier decisions become clinically consequential.

The sixth factor is proportionality. Once multiple actors are identified as contributors, responsibility should be allocated according to their relative contribution to the relevant risk. Proportionality prevents two opposite errors: concentrating all liability on the physician and imposing undifferentiated liability on every participant. A physician who ignores a clear warning may bear substantial responsibility even if the AI system has minor imperfections. Conversely, a physician who reasonably relies on a seriously defective system may bear little or no responsibility while the developer and hospital carry greater responsibility.

Proportionality should consider the seriousness of the breach, degree of control, foreseeability, available safeguards, causal contribution, and economic or organizational capacity to prevent recurrence. Such an approach also supports efficient prevention because actors are encouraged to invest in safeguards where they have the greatest ability to reduce risk. Liability thus becomes not merely compensatory but also regulatory and preventive.

On this basis, the proposed Indonesian framework can be summarized as shown Table 1.

TABLE 1. Summary of Potential Liability on AI-related Injuries

Actor	Primary Sphere of Control	Relevant Duty	Typical Breach	Potential Liability
Physician	Clinical decision	Reasonable clinical judgment and informed use of AI	Blind reliance, misuse, ignoring contradictory evidence	Professional and civil liability; criminal liability only where legally justified
Hospital	Procurement, deployment, training, monitoring	Safe institutional governance	Inadequate validation, training, supervision, monitoring	Institutional/civil liability
AI Developer	Design, training, validation, software	Technological safety and disclosure	Defective design, inadequate validation, misleading claims, undisclosed limitations	Product/civil liability
Data Controller	Collection and processing of health data	Lawful and secure data governance	Unauthorized processing, serious data-quality failures	Data/privacy liability
Integrator/ Vendor	Technical implementation	Correct integration and configuration	Faulty implementation or unauthorized modification	Contractual/civil liability
Multiple Actors	Interdependent risks	Coordinated safety management	Cumulative or concurrent failures	Concurrent/proportionate liability

Source: Various sources, analyzed by the Authors

This framework (Table 1) should also distinguish between civil, professional disciplinary, and criminal responsibility. Civil liability is the most suitable mechanism for complex AI-related injuries because it can address compensation and allocate loss among multiple actors without requiring proof of criminal intent or an exceptionally high degree of individual culpability. Professional disciplinary

mechanisms are appropriate where a physician's conduct violates professional standards even if the conduct does not justify civil or criminal sanctions. Criminal law should remain exceptional because the complexity of AI causation makes it particularly dangerous to equate an adverse outcome with criminal wrongdoing. A physician should not face criminal prosecution simply because an AI system produced a false diagnosis. Criminal responsibility should require the elements of the relevant offence and the legally required degree of fault. This differentiation is essential to prevent the criminal law from discouraging responsible adoption of beneficial technologies. It also protects the integrity of professional medicine by distinguishing unavoidable clinical uncertainty from reckless or seriously culpable conduct.

The proposed model also has implications for the burden of proof. Patients ordinarily face significant informational disadvantages because they do not control the algorithm, its training data, its validation records, or the hospital's technical logs. Requiring patients to prove the precise internal operation of a proprietary AI system may effectively make legitimate claims impossible. The law should therefore consider carefully structured disclosure and evidentiary presumptions. Where a claimant establishes that an AI system was used in the relevant clinical decision, that the system produced the relevant output, that the output materially contributed to the treatment decision, and that the resulting harm is consistent with a known category of system risk, defendants should be required to produce relevant records concerning validation, warnings, updates, and system performance. This would not create automatic liability. Rather, it would correct the informational imbalance that otherwise allows technologically complex defendants to avoid accountability simply because the evidence is inaccessible to patients. The principle is consistent with the broader WHO emphasis on accountability and mechanisms for individuals adversely affected by algorithmic decisions to challenge those decisions and obtain redress (World Health Organization 2021).

Data governance is also inseparable from this liability model. Law Number 27 of 2022 concerning Personal Data Protection establishes a broader framework for the processing and protection of personal data, while health information is treated as a particularly sensitive category. AI systems depend upon extensive data collection, processing, storage, and potentially cross-organizational transfer. A diagnostic system may therefore create both clinical and informational risks. If patient data are processed unlawfully, inadequately secured, or used beyond the authorized purpose, the resulting legal responsibility may arise independently of whether the AI diagnosis was correct. The distinction is important because a patient could suffer two different forms of harm: physical harm from a diagnostic error and privacy harm from improper data processing. These should not be collapsed into one liability category. The distributed framework instead recognizes that different actors may owe different duties concerning data, clinical care, and technology. Healthcare institutions may be controllers or users of patient information, while

developers may process data under contractual or technical arrangements. Responsibility should correspond to the role each actor actually performs.

The electronic medical-record framework further supports this approach. Regulation of the Minister of Health Number 24 of 2022 establishes the Indonesian regulatory framework for medical records and provides a foundation for electronic documentation. In AI-assisted healthcare, documentation should evolve beyond the traditional record of human observations and final diagnoses. For clinically material AI decisions, the system version, relevant output, warnings, and human response should be sufficiently documented to permit later reconstruction. This is not merely a litigation requirement. It is a patient-safety requirement because institutions need to identify patterns of algorithmic failure. If several patients experience similar false negatives, the hospital should be able to determine whether the problem is associated with a particular model version or implementation setting. Without auditability, neither quality improvement nor legal accountability can function effectively. The law should therefore treat AI audit trails as part of the broader integrity of clinical documentation.

Comparative regulation supports the lifecycle dimension of this framework. The European Union's Product Liability Directive 2024/2853 is particularly instructive because it explicitly adapts product-liability rules to digital technologies and software, including AI systems, and recognizes continuing manufacturer responsibility in circumstances involving software updates, upgrades, and machine-learning algorithms under manufacturer control. The importance of this development for Indonesia lies not in direct transplantation but in its recognition that digital products challenge the assumption that liability can be fixed at the moment of sale. AI may continue to evolve or be modified after deployment. Consequently, legal responsibility should track continuing control. A developer that retains the ability to update the system should retain corresponding safety obligations. Similarly, a hospital that modifies or configures the system substantially should assume responsibility for risks created by its intervention. This creates a dynamic allocation of liability consistent with the technological reality of AI.

The proposed framework can therefore be conceptualized as a risk-control matrix. At the design stage, developers bear primary responsibility because they control architecture, training, testing, and technical specifications. At the validation stage, developers and deploying institutions share responsibility because technical performance must be assessed against the actual clinical context. At procurement, the hospital assumes greater responsibility because it decides whether the technology is appropriate for its environment. At deployment, hospitals and physicians share responsibility for safe integration and reasonable clinical use. At monitoring and updating, responsibility becomes distributed again according to who controls the relevant system changes. This lifecycle model is preferable to a static liability framework because AI risk changes over time. It also creates incentives for every actor to maintain safety rather than treating responsibility as

something that can be transferred contractually or administratively to another participant.

Contracts between hospitals and AI developers should not be permitted to undermine patient-protection principles. A hospital may agree with a developer that the developer will not be responsible for certain categories of loss, but contractual allocation between commercial parties should not necessarily determine the patient's legal rights. Otherwise, sophisticated technology providers and healthcare institutions could allocate liability privately in ways that leave patients without effective remedies. Contractual provisions can determine contribution and indemnification between the parties, but public law should preserve the patient's ability to seek compensation from actors whose legally relevant conduct caused harm. This distinction is especially important because patients generally do not participate in the hospital-developer contract. They should not bear the consequences of contractual arrangements to which they were not parties. The distributed framework therefore concerns substantive responsibility to patients, while private contractual arrangements may subsequently redistribute financial responsibility between commercial actors.

The framework also has an innovation dimension. Excessively broad liability can discourage technological development, but inadequate liability can encourage irresponsible commercialization. The appropriate regulatory objective is therefore not maximum liability but optimal accountability. Developers should have incentives to validate systems rigorously, disclose limitations, monitor performance, and correct defects. Hospitals should have incentives to conduct procurement due diligence, train staff, monitor outcomes, and suspend unsafe technologies. Physicians should have incentives to maintain professional competence and avoid blind reliance on automated recommendations. Patients should have access to effective redress when these safeguards fail. The balance is therefore not between innovation and regulation as mutually exclusive objectives. Responsible liability can support innovation by creating predictable rules for risk allocation. When developers know the conditions under which they may be liable, they can incorporate those risks into insurance, contracts, product design, and compliance systems. Legal certainty may consequently encourage rather than undermine responsible technological investment.

Finally, Indonesia should consider developing a dedicated regulatory framework for clinical AI that does not merely regulate AI generally but identifies healthcare-specific risks. Such a framework should define high-risk clinical AI applications, establish minimum validation requirements, require documentation of intended use and limitations, provide standards for human oversight, mandate incident reporting, establish post-deployment monitoring, and clarify allocation of responsibility among developers, institutions, and users. The framework should also establish procedures for preserving and disclosing AI-related evidence in patient injury disputes. The WHO principles provide a useful normative foundation, while comparative regulatory developments demonstrate that risk-based and lifecycle-

oriented governance is increasingly feasible (World Health Organization 2021). Indonesia does not need to reproduce the European Union's regulatory architecture. Instead, it should construct a framework consistent with Indonesian health law and institutional capacity while adopting the underlying principle that technological risk must remain connected to identifiable human and organizational responsibility.

Conclusion

AI-assisted diagnosis exposes a fundamental limitation in conventional medical liability law because the clinical decision is increasingly produced through interaction between human judgment and technological systems. Indonesian Law Number 17 of 2023 provides important foundations for addressing this transformation by maintaining professional responsibility for medical personnel while recognizing legal responsibility at the hospital level Nevertheless, the legislation does not expressly establish a comprehensive regime for algorithmic medical liability. The resulting gap should not be filled by automatically assigning responsibility to physicians whenever AI contributes to an adverse outcome. Such an approach would confuse technological error with professional negligence and could transform physicians into insurers against risks that they did not create and could not reasonably control. At the same time, treating AI developers as universally liable would be equally inappropriate because AI systems are probabilistic and clinical outcomes depend on multiple human and institutional decisions. The appropriate solution is a distributed framework in which physician, hospital, and developer responsibility is determined by human control, risk creation, foreseeability, transparency, breach of duty, and causal contribution. Under this model, professional responsibility remains central but is no longer treated as synonymous with technological responsibility.

The proposed reconstruction positions AI medical liability as a lifecycle and multi-actor problem. Physicians should remain responsible for unreasonable clinical reliance, misuse, failure to exercise professional judgment, or disregard of clinically significant evidence. Hospitals should assume responsibility for technology procurement, validation, training, implementation, supervision, monitoring, and institutional governance. Developers should be accountable for defective design, inadequate validation, materially misleading claims, undisclosed limitations, unsafe updates, and other technological risks within their sphere of control. Where multiple breaches materially contribute to patient injury, concurrent and proportionate liability should be available rather than forcing the law to identify a single defendant. This approach should be supported by stronger AI audit trails within electronic medical records, meaningful disclosure obligations, appropriate evidence-preservation requirements, and lifecycle monitoring. The ultimate objective should not be to determine whether humans or machines are “to blame,” but to construct a legal system in which responsibility follows the actor who created

or controlled the relevant risk. Such a model can strengthen patient protection, preserve professional fairness, and create predictable incentives for responsible AI innovation in Indonesian healthcare.

References

European Union. 2024. "Directive (EU) 2024/2853 of the European Parliament and of the Council of 23 October 2024 on Liability for Defective Products and Repealing Council Directive 85/374/EEC." *Official Journal of the European Union*.

Gerke, Sara, Timo Minssen, and I. Glenn Cohen. 2020. "Ethical and Legal Challenges of Artificial Intelligence-Driven Healthcare." In *Artificial Intelligence in Healthcare*, edited by Adam Bohr and Kaveh Memarzadeh, 295–336. London: Academic Press. https://doi.org/10.1016/B978-0-12-818438-7.00012-5.

Ghassemi, Marzyeh, Luke Oakden-Rayner, and Andrew L. Beam. 2021. "The False Hope of Current Approaches to Explainable Artificial Intelligence in Health Care." *The Lancet Digital Health* 3 (11): e745–e750. https://doi.org/10.1016/S2589-7500(21)00208-9.

Holzinger, Andreas, Andrй Carrington, Heimo Мьller, et al. 2020. "Explainability for Artificial Intelligence in Healthcare: A Multidisciplinary Perspective." *BMC Medical Informatics and Decision Making* 20. https://doi.org/10.1186/s12911-020-01332-6.

Indonesia. 2022. *Law Number 27 of 2022 concerning Personal Data Protection*.

Indonesia. 2022. *Regulation of the Minister of Health Number 24 of 2022 concerning Medical Records*.

Indonesia. 2023. *Law Number 17 of 2023 concerning Health*. State Gazette of the Republic of Indonesia 2023, No. 105.

Indonesia. 2024. *Government Regulation Number 28 of 2024 concerning Implementing Regulations of Law Number 17 of 2023 concerning Health*. State Gazette of the Republic of Indonesia 2024, No. 135.

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.

Markus, Aniek F., Jan A. Kors, and Peter R. Rijnbeek. 2021. "The Role of Explainability in Creating Trustworthy Artificial Intelligence for Health Care: A Comprehensive Survey of the Terminology, Design Choices, and Evaluation Strategies." *Journal of Biomedical Informatics* 113: 103655. https://doi.org/10.1016/j.jbi.2020.103655.

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.

Price, W. Nicholson II, Sara Gerke, and I. Glenn Cohen. 2019. "Potential Liability for Physicians Using Artificial Intelligence." *JAMA* 322 (18): 1765–1766. https://doi.org/10.1001/jama.2019.15064.

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

***

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