

Article

Research on Liability Allocation Mechanisms for Harm Caused by Medical Artificial Intelligence under the Civil Law Framework

Jiaran Zhang^{1,*}

¹ Dalian Ocean University, Dalian, China

* Correspondence: Jiaran Zhang, Dalian Ocean University, Dalian, China

Abstract: Medical artificial intelligence distributes clinical decision-making among doctors, hospitals, software companies, manufacturers, system integrators, and data and infrastructure providers. When harm occurs, this distributed responsibility complicates traditional civil-law determinations of fault, product defect, causation, disclosure, and proof. Drawing on foreign legal frameworks and China's existing civil-liability rules, this paper examines the principles of civil liability and product safety legislation in China. It believes that neither the exclusive liability of clinicians nor general enterprise liability is suitable for handling the risks of medical AI. Patients may pursue claims under professional negligence, organisational fault, breach of contract, or product liability. Liability should then be allocated according to operational control, preventability, risk creation, breach of professional or technical duty, evidentiary failure, and causal contribution. A structured formula should guide apportionment, without being fully automated. Rebuttable evidentiary presumptions, mandatory versioned logging, proportionate disclosure, independent auditing, post-deployment monitoring, risk-sensitive insurance, etc., can all reduce information asymmetry and protect legitimate rights and interests. With good calibration, joint and several liability can provide compensation and accountability for harm without labelling every unfortunate event as misconduct and thus prevent stifling good innovation.

Keywords: medical artificial intelligence; civil liability; product liability; causation; risk allocation

1. Introduction

A patient may be harmed because a doctor followed the diagnosis or treatment advice provided by AI. However, the fundamental cause may lie elsewhere and predate the clinical encounter. The training data may not have been representative of the patient group; the validation was insufficient; there may have been an error in the institutional interface; a software update could have caused drift; or the clinician may have accepted an unreasonable result without proper confirmation. Traditional malpractice rules focus on the conduct of doctors, while product liability holds manufacturers responsible for defects. Neither approach is sufficient to address the distributed nature of harm [1].

Medical AI encompasses software that employs statistical and machine-learning techniques to assist in diagnosis, prognosis, triage, treatment, and monitoring. Its autonomy varies considerably: assistive tools depend on expert interpretation, whereas highly autonomous systems operate with minimal real-time human intervention. This spectrum of autonomy directly shapes key legal constructs—including reasonable reliance, warning duties, and the locus of effective control. Accordingly, regulatory classifications must accurately reflect operational reality rather than obscure the system's actual function.

Compensatory civil liability is covered in this paper; criminal, disciplinary and administrative liability are not. At its core is the idea that liability should be in line with legal responsibility for control, prevention of harm, creation of risk, professional/technical

Received: 28 June 2026

Revised: 07 August 2026

Accepted: 18 August 2026

Published: 24 August 2026



Copyright: © 2026 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

obligations, disclosure failures and causal contribution [2]. Regulatory compliance is relevant, but it does not automatically establish fault or negligence. The analytical framework proposed in this paper outlines the grounds for a claim and specifies how responsibility should be divided among the parties [3].

2. Civil-Law Foundations and Distinctive Risks

2.1. Fault, Defect, Damage and Attribution

Generally speaking, professional negligence consists of a duty of care, a breach, damage and causation. Deficiencies in procurement, staffing, training, maintenance, and management are borne by healthcare institutions. A contract binds parties to provide services and receive payment. Product liability essentially examines whether software or a software-enabled device conforms to mandated safety standards. The updated Product Liability Directive in the EU now covers software, updates, cybersecurity risks and the lack of evidence for complex technical products, and the regulations for medical devices have separate stages of risk management, clinical evaluation, surveillance and corrective action [4].

As shown in Table 1, the routes are not disjoint but rather overlap. An erroneous warning may not be acted upon, yet the situation remains unchanged. Courts should not require the claimant to identify a single theory of liability before ordering disclosure of records that are solely within the defendants' control.

Table 1. Conventional Civil-Liability Routes Applied to Medical AI

Liability route	Claimant's core burden	Principal defendant	Typical defense	AI-specific difficulty
Professional fault	Duty, breach, harm, causation	Clinician	Reasonable professional judgment	Output opacity may conceal technical error
Organizational fault	Institutional breach and causation	Healthcare institution	Reasonable procurement and governance	Responsibility is divided across departments
Product liability	Defect, damage, causal link	Manufacturer or producer	No defect when placed in circulation; statutory defense	Model versions and updates change product state
Contract	Breach, loss, contractual nexus	Supplier or deployer	Compliance with agreed specification	Patients may lack contractual privity
Vicarious liability	Employee fault within assigned functions	Institution or employer	Actor outside relevant relationship	Human and machine conduct may be intertwined

2.2. Opacity, Adaptation, Data Dependence, and Automation Bias

Opacity is problematic because an injured patient can only see the output and the adverse event, not the computational path that led to it. Opacity is not necessarily a defect; some excellent systems are too complex to understand, and an explainable system can still be very safe. Legally, opacity affects the adequacy of verification, warnings, contestability, and proof.

Adaptation has a problem of time: which model, dataset, threshold and configuration were used at the time of recommendation? Data dependence also creates population-specific risk. A model validated in one clinic will change with the increase in the number of patients, machines, workflow, and so on. Automation bias is also a reason why people think the output is correct. Therefore, recent medico-legal research has begun to carry out life-cycle assessments and is no longer focused only on the final clinician-patient interaction (As shown in Figure 1).



Figure 1. Chain of medical-AI Risk from Development to Patient Harm; Author's Normative Synthesis, Not an Empirical Dataset.

3. Liability of Principal Actors

3.1. Healthcare Institutions and Professionals

Clinicians should make recommendations based on the patient's individual circumstances, available options, and relevant professional norms. Clinicians should check the output against obvious contraindications and report any anomalies. However, it is unreasonable to expect clinicians to independently detect latent defects in training data or source code. Therefore, reasonable reliance should be based on the authorised purpose of the system, actual operation, warnings, local conditions, urgency and the professional's reasonable capacity to intervene [5].

Institutions must organise procurement and workflow. Their responsibilities are to set goals, supervise local operations, build relationships, train staff, guarantee network security, establish human-in-the-loop rules, handle emergencies, etc. Deployment outside the validated population or without a proper escalation path should be regarded as an organisational defect, even if no single clinician behaved unreasonably. A competent regulatory authority can also protect patients from harm caused by unforeseeable latent defects in products.

Informed consent should include the main reasons for using AI in healthcare and how it will affect a reasonable person's choice. Disclosure need not mention technical architecture. Present the goals of the system, the degree of human intervention, known deficiencies and reasonable clinical choices. Consent does not relieve one of the responsibility for negligence or product liability; thus, if there has been a failure to disclose, liability will only be incurred if the omitted information was material and legally linked to the damage [6].

3.2. Developers, Manufacturers, and Deployers

The reasons for the problem of producer responsibility may include a poor architecture, non-representative development data, lack of validation, an unsafe interface design, absence of warnings, cybersecurity deficiencies, harmful updates or ineffective post-market surveillance, etc. Artificial intelligence in medical devices in the EU is considered a high-risk type of artificial intelligence and must meet many requirements for risk management, data governance, documentation, logging, transparency, human supervision, accuracy, robustness and post-market monitoring under the Artificial Intelligence Act [7].

Gerke and Simon have found that new litigation can put forward liability theories for manufacturers based on representations, design choices and foreseeable clinical reliance, and marketing claims are a typical case; a supplier who advertises autonomous-quality performance should not ignore all clinically insignificant results. The deployer and the integrator will be responsible for any serious modifications to the system, incorrect connection of data, ignoring warnings, and setting thresholds without verification.

3.3. Data and Infrastructure Providers

Not all the people in the AI development and application chain are responsible for the harm caused to end-users. Liability should be shared among the various parties in proportion to their roles, responsibilities and actual behaviour. If the dataset provided by the data supplier is seriously falsified or misrepresented, or if it does not meet the required data quality standards in advance, then the data supplier shall bear the legal liability for any resulting damages. Cloud or platform service providers will bear the legal responsibility for negligent service disruptions, a lack of security in the underlying technical infrastructure, unauthorised modification of the components of the AI system, or serious violations of statutory and contractual supervision obligations. Contractual

warranties and indemnity clauses can distribute and reduce the business risks among the parties to the contract reasonably, but they cannot override or contradict mandatory provisions of the law protecting the rights of patients. Any contractual clause that purports to exclude or limit the aforementioned non-waivable patient-oriented guarantees may be challenged and is likely to be held invalid by a court, in accordance with mandatory provisions of applicable law.

Table 2 is qualitative: it connects liability to controllable risk instead of assigning automatic responsibility to every participant.

Table 2. Actor-specific Controls, Duties, Evidence, and Defenses

Actor	Controllable risk and duty	Key evidence	Potential basis	Legitimate defense
Clinician	Contextual review and escalation	Clinical notes; override record	Professional fault	Reasonable reliance
Institution	Procurement, training, integration	Policies; local validation	Organizational liability	Sound governance
Developer/manufacturer	Design, validation, warning, update	Technical file; test results	Product liability or negligence	No defect or causal effect
Deployer/integrator	Configuration and workflow	Configuration history	Negligence or contract	Authorized implementation
Data provider	Provenance and representations	Dataset documentation	Negligence or contract	Accurate disclosure
Infrastructure provider	Availability and security	Audit and incident logs	Negligence or contract	External, unforeseeable cause

4. Causation and Evidence

4.1. Multiple and Cumulative Causes

Disease progression, clinician's judgement, data quality, interface design and model errors can all cause harm. Factual causation refers to whether the relevant breach or defect caused the legal consequence; the rules on the scope of liability determine whether the resulting damage falls within the risk that made the behaviour unlawful. Loss-of-a-chance may be recognised to increase the likelihood of recovery in some jurisdictions, but not in others. According to the above evidence, classify the injuries. When there are many reasons for an indivisible injury, the rules of solidarity or joint liability should provide compensation and postpone the division of fault [8].

4.2. Information Asymmetry and Proof

Patients seldom have the model version, input record, confidence score, validation report, update history or internal incident analysis. Ordinary burdens of proof will be too heavy to meet before disclosure takes place. The appropriate response is to require preservation and disclosure. If the claimant believes that the use of AI is related to the injury, the court should order the production of relevant records.

An actor who controls a legally required record but fails to preserve or disclose it reasonably should face a presumption that the non-preservation or non-disclosure was unjustified. The presumption should be applied on a case-by-case basis. The defendant may offer proof of a cause of action, reasonable notice, unforeseeable circumstances, or that the disputed output did not affect the treatment in response to this [9].

Table 3 proposes procedural consequences rather than reporting litigation statistics.

Table 3. Evidence Required Along the Causal Chain

Disputed issue	Evidence holder	Required record	Probative function	Unjustified nonproduction
Input accuracy	Institution	Timestamped patient inputs	Tests data-to-output connection	Presumption of material inaccuracy
Model identity	Manufacturer /deployer	Version and update log	Identifies relevant product state	Presumption favoring alleged version
Known limitations	Manufacturer	Validation and risk records	Tests defect and warning adequacy	Adverse inference on knowledge
Clinical reliance	Institution/clinician	Output and override log	Shows influence on treatment	Presumption of material reliance
Post-market response	Manufacturer	Incident and monitoring file	Tests preventability after notice	Inference of delayed response

5. A Workable Allocation Mechanism

5.1. Entitlement Before Contribution

Allocation should proceed in two stages. Firstly, there should be compensable harm and a qualifying link to the use of medical AI in the relevant cause of action. The above regulations set out rules for the liability of service providers towards patients. Second, the responsible parties distribute the share. This two-stage approach avoids increasing the complexity of supply-chain apportionment and allows for a more equitable division of contributions.

5.2. Allocation Formula

For responsible actor i , the suggested share is:

$$L_i = D \times \frac{W_i}{\sum_{j=1}^n W_j} \quad \# \quad (1)$$

where

$$W_i = \alpha C_i + \beta P_i + \gamma R_i + \delta B_i + \varepsilon E_i + \zeta K_i \quad \# \quad (2)$$

Here, L_i is actor i 's allocated amount; D is legally recoverable damage; C_i is operational control; P_i is capacity to prevent harm; R_i is risk creation; B_i is breach of professional or technical duty; E_i is evidentiary or disclosure failure; and K_i is causal contribution. Coefficients α through ζ are jurisdiction-sensitive weights.

The inputs are reasoned ordinal judgments, not observed probabilities. Courts should explain the scores with reference to evidence and apply the same scale to all parties. There is a formula for comparison and settlement, but mandatory product liability rules, rules on joint liability, contribution statutes, claimant fault, insolvency and causal thresholds must not be ignored [10].

Table 4 is the author's normative framework. It disciplines judgment but does not conceal value choices behind numerical appearance.

Table 4. Proposed Allocation Factors and Safeguards

Factor	Judicial question	Indicative evidence	Reason for inclusion	Double-counting safeguard
Control	Who could alter the risky process?	Access and configuration rights	Responsibility should follow power	Separate control from breach
Preventability	Who could efficiently avert injury?	Feasible safeguards and notice	Encourages precaution	Consider only ex ante capacity

Risk creation	Who introduced the relevant danger?	Design and deployment records	Connects enterprise activity to loss	Identify the specific risk
Duty breach	Which legal obligation was violated?	Standards, policies, warnings	Preserves doctrinal foundation	Do not rescore causal magnitude
Evidence failure	Who impaired reliable adjudication?	Retention and disclosure history	Corrects information asymmetry	Apply only to missing material evidence
Causal contribution	How strongly did conduct affect harm?	Logs, clinical evidence, expert analysis	Differentiates concurrent causes	Avoid repeating risk creation

5.3. Decision Sequence and Defenses

Figure 2 ensures that apportionment occurs only after the legal basis of recovery is established. Available defenses may include a statutory development-risk defense, reasonably unforeseeable misuse, adequate warning, contributory fault, or learned-intermediary reasoning where the jurisdiction recognizes it. Regulatory compliance is relevant evidence of care and expected safety, but minimum compliance should not be conclusive when the producer knew of a serious residual risk or promoted use outside validated conditions.

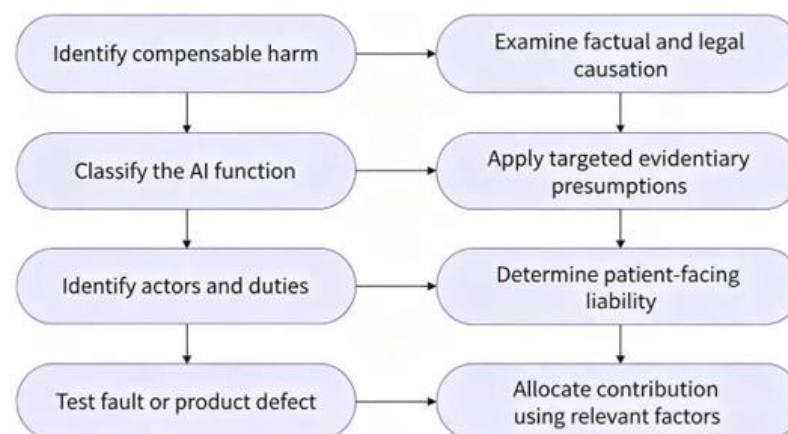


Figure 2. Author-developed Decision Process for medical-AI Liability Allocation.

6. Practical Safeguards and Implementation

Lifecycle governance prevents damage and organises evidence collection. Versioned logs should record model information, input, output, confidence scores, modifications, and manual changes, while maintaining privacy and reasonableness. Dataset provenance, subgroup validation, incident reporting, scheduled audits, cybersecurity tests, drift detection, etc., can all help us find changes at a particular time. According to the WHO's basic principles of governance, these systems should be open, accountable, inclusive, and humane. The US Food and Drug Administration has added the regulation of AI-powered medical devices to the list of problems [11].

Consent should be staged; present only the essential information on the first occasion and provide more details upon request. Institutions have particular training and appeal channels. Commercial contracts should stipulate audit, update, notification and indemnity obligations, but not limit patients' remedies. High-risk use cases should be covered by insurance. A small-scale compensation fund could cover residual harm caused by AI where no specific breach of duty by any individual can be proven.

As Table 5 shows, evidence infrastructure is not merely defensive documentation: it is part of safe system design.

Table 5. Implementation Measures and Legal Effects

Measure	Responsible actor	Prevention function	Evidentiary function	Civil-liability effect	Status
Versioned logging	Manufacturer and institution	Detects errors and drift	Reconstructs event	Supports causation analysis	Regulatory basis plus proposal
Dataset provenance	Developer	Reveals limitations	Identifies data source	Informs defect inquiry	Regulatory basis
Local validation	Institution	Tests contextual safety	Documents reasonable deployment	Relevant to organizational fault	Normative recommendation
Human-override record	Institution	Exposes automation bias	Shows clinical reliance	Informs professional fault	Normative recommendation
Risk-tiered insurance	Producer and deployer	Funds risk management	Confirms coverage	Improves compensation	Normative recommendation

7. Conclusion

Medical AI does not make established civil-law categories obsolete. It also reveals the shortcomings of relying exclusively on these traditional categories. Professional negligence is the lack of good clinical judgment; organisational liability arises from procurement and workflow defects; product liability refers to unsafe software, warnings and updates, etc. The challenge is to integrate these routes within a distributed-control chain.

A two-step model should be used: first, determine the patient's entitlement; second, avoid prematurely assuming exclusive technical causes. Attribution should then be based on control, preventability, risk creation, breach, evidentiary conduct, and causal contribution. Targeted presumptions can reduce information asymmetry without taking full responsibility for all the negative consequences, and defendants can still present some rebuttals and other defences. These are inevitable uncertainties in complex systems, not necessarily deficiencies in technology or professional practice.

Trustworthy innovation does not have to be explainable. It should have traceable versions, representative validations, usable warnings, accountable human supervision, monitoring and insurance that can turn responsibility into compensation. Civil liability should do its job by disclosing foreseeable risks, have the party with more control bear the cost of such risks, and leave space for good systems in which the remaining risks have been honestly disclosed and responsibly managed.

References

1. X. Shentu, "A review on legal issues of medical robots," *Medicine*, vol. 103, no. 21, p. e38330, 2024.
2. D. Rimkutė, "AI and liability in medicine: The case of assistive-diagnostic AI," *Baltic Journal of Law & Politics*, vol. 17, no. 2, pp. 64–81, 2024.
3. P. Nolan and R. Matulionyte, "Artificial intelligence in medicine: Issues when determining negligence," *Journal of Law and Medicine*, vol. 30, no. 3, pp. 593–615, 2023.
4. C. Giorgetti, A. Giorgetti, and R. Boscolo-Berto, "Establishing new boundaries for medical liability: The role of AI as a decision-maker," *Advances in Clinical and Experimental Medicine*, vol. 34, no. 10, pp. 1601–1606, 2025.
5. S. Sankari et al., "Liability for Medical Devices - How Clear and Harmonised Is It?," *Europarättslig tidskrift*, no. 4, pp. 651–668, 2023.
6. Y. Wang and Z. Zhou, "Medical damage liability risk of medical AI: from the perspective of DeepSeek's large-scale deployment in Chinese hospitals," *Frontiers in Public Health*, vol. 13, p. 1726205, 2025.
7. S. Wolfhagen, "Dr Robot, Defendant: Medical Negligence Claims in a Post-AI World," *Precedent*, no. 183, p. 16, 2024.

8. M. Geny et al., "Liability of health professionals using sensors, telemedicine and artificial intelligence for remote healthcare," *Sensors*, vol. 24, no. 11, p. 3491, 2024.
9. M. Herrmann, A. Wabro, and E. Winkler, "Percentages and reasons: AI explainability and ultimate human responsibility within the medical field," *Ethics and Information Technology*, vol. 26, p. 26, 2024.
10. C. Zhang, "The legal dilemma of medical artificial intelligence in China: challenges to physicians' duty to inform and a typology-based response," *Frontiers in Public Health*, vol. 14, 2026.
11. M. M. Mello and N. Guha, "Understanding Liability Risk from Using Health Care Artificial Intelligence Tools," *New England Journal of Medicine*, vol. 390, no. 3, pp. 271–278, 2024.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of Publisher and/or the editor(s). Publisher and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.