Asian Journal of Law and Society (2026), **13**, 630–653
doi:10.1017/als.2026.10070

CAMBRIDGE UNIVERSITY PRESS

RESEARCH ARTICLE

Establishing Responsibility Systems for AI Application in Clinical Decision-Making: A Relational Responsibility Perspective

Xiyi Chen

China Institute for Socio-Legal Studies, KoGuan School of Law, Shanghai Jiao Tong University, China
Email: chenxiyi@sjtu.edu.cn

Abstract

AI in clinical decision-making challenges liability frameworks grounded in subject–object dichotomies, linear causation, and atomised attribution. Drawing on relational responsibility, this article argues that the human–machine association, rather than the AI system or clinician alone, should be treated as the fundamental decision-making unit. Responsibility should therefore emphasise interaction quality, harm prevention, and relationships that sustain future cooperation. The article proposes dialogic responsibilities for technology providers, medical institutions, clinicians, and patients. For external compensation, medical institutions should act as the primary responsibility hub, reducing patients' burden of proof while retaining rights of internal recourse based on specific human–machine interaction patterns. Internally, liability should reflect calibrated trust, documented dialogue, and the respective commitments of technology providers and clinical users. It thereby promotes accountable collaboration throughout the clinical AI lifecycle. By linking physician–patient trust with algorithmic trust, this relational framework seeks to reconcile technological innovation, professional autonomy, and the protection of patients' rights.

Keywords: clinical decision-making AI systems; human–machine association; relational responsibility; trust relationships; dialogic responsibility

1. Problem statement: the dilemma of liability systems in the application of clinical decision-making AI systems

Clinical decision-making AI systems analyse complex medical data (e.g., medical imaging, electronic health records, laboratory results, genomic data) to generate specific, actionable clinical recommendations. Such systems may operate based on predefined rules and information or leverage machine learning (particularly deep learning) to provide diagnostic and treatment support for healthcare professionals (*see* Grasso, 2024, p. 155). According to the National Health Commission of the People's Republic of China (hereinafter "NHC"), the NHC has launched pilot programmes for medical AI-assisted diagnosis and treatment systems in 27 counties (cities, districts) across 9 provinces, including Jiangsu, Shandong, and Guangdong. These systems assist physicians in diagnosing and treating common and frequently occurring diseases. By 2024, primary healthcare institutions in the project implementation areas had provided nearly 40,000

© The Author(s), 2026. Published by Cambridge University Press in association with Shanghai Jiao Tong University. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 16 Nov 2026 at 18:59:47, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/als.2026.10070

instances of diagnostic recommendations through the AI-assisted systems (*see* National Health Commission of the People's Republic of China, 2025).

This article focuses on clinical decision-making AI systems because such systems serve as a prism reflecting the interactions among multiple stakeholders—including developers, manufacturers, distributors, healthcare institutions, medical professionals, and patients—and clearly reveal the institutional demands generated by emerging practices. Specifically, first, these systems most vividly demonstrate the impact of cutting-edge AI—rooted in deep learning and large language models with emergent properties—on medical practice. Their algorithmic black-box nature also highlights the tension between technological innovation and healthcare safety. Second, these systems directly engage with core medical functions. Unlike data preprocessing systems and image recognition and processing systems that serve as prerequisites for clinical decision-making, intelligent administrative systems for optimising medical workflows and management, or surgical robots executing specific procedures, clinical decision-making AI systems more directly reshape traditional models of clinical practice and physician–patient relationships (*see* Liang, 2023). This distinction is also reflected in China's National Medical Products Administration (NMPA) Guidelines for Classifying Artificial Intelligence Medical Software Products and the Medical Device Classification catalogue's *Medical Software* sub-catalogue issued by NMPA's Center for Medical Device Evaluation in 2021. This catalogue categorises five types of software: treatment planning, image processing, data processing, decision support, and in vitro diagnostics. Software involving image processing and data processing is subject to Class II management standards, while software involving clinical decision-making mostly follows the stricter Class III management standards (*see* Center for Medical Device Standardization Administration NMPA, no date).

In light of China's urgent demand for diagnostic resources and technological innovation, coupled with high expectations for the responsibilities borne by medical institutions, and the simultaneous benefits and risks in AI system applications, the question of how legal systems can establish responsible parties and promote accountable actions and interactions among all parties warrants urgent attention. This inquiry is essential to regulating the risks associated with these systems and ensuring clear allocation of responsibility when actual harm occurs. Within China's current legal framework, this question has generally been framed in the literature as "the respective applicability of product liability and medical malpractice liability," further decomposed into three key disputes: first, the characterisation of clinical decision-making AI systems—specifically, whether they can be classified as products and thus subject to product liability regulations. Second, the functional positioning of such systems and the corresponding degrees of duty of care that should be borne by various parties, including technology providers, sellers, medical institutions, and healthcare professionals. This involves regulating the entire process from system development to deployment. Third, the specific model for assigning liability after harm occurs, with the core issue being how liable parties should compensate resulting harm. To address these issues, this article first reviews existing research, reflecting on its insights and limitations. It then reexamines the characteristics of clinical decision-making AI systems and their practical applications, proposing the concept of *relational responsibility* from a jurisprudential perspective to provide a novel entry point for constructing specific institutional arrangements. Finally, it moves beyond the limitations of existing research by establishing trust relationships as the value foundation and developing detailed institutional proposals for external and internal responsibility arrangements.

2. The challenge of balancing value tensions and the failure of the binary analytical framework

First, regarding the classification of clinical decision-making AI systems, existing research largely follows a product-service dichotomy framework. The classification is based

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 16 Nov 2026 at 18:59:47, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/als.2026.10070

primarily on three considerations: (1) whether the delivered content possesses tangible physical attributes; (2) whether the user's actions carry final determinacy, thereby rendering independent discussion of product liability unnecessary; and (3) whether the corresponding liability burdens imposed on producers and sellers under different classifications are reasonable. The product theory represents the prevailing academic consensus in China, as reflected in regulatory policies, such as the preceding Guidelines for Classifying Artificial Intelligence Medical Software Products. Its proponents primarily offer three arguments: (1) The concept of product is sufficiently flexible, and legal systems should adapt to the digital era by incorporating intangible software into the product category. (2) While medical professionals retain decisive authority over clinical judgements, this does not preclude acknowledging the critical role of AI systems in the process. This recognition justifies holding technology providers accountable. To fully protect patient rights and avoid undue burdens on medical institutions and practitioners, product liability frameworks should be established to broaden the scope of responsible parties (*see* Hou, 2025, pp. 66–67; Zheng, 2023, p. 215). (3) Even when AI systems play a significant role in diagnosis and treatment, institutional design should emphasise human dominance and irreplaceability through classifying systems as products rather than autonomous decision-makers (*see* Li, 2020, p. 54). It is evident that the definition of product itself is not the core issue here. Classifying these systems as products primarily stems from functional considerations, aiming to include all actors influencing decision-making within the scope of liability while avoiding unduly stringent accountability that could hinder practical application. However, opposing views to the product theory argue that the deep learning capabilities of these systems render them beyond the control of technology providers. Consequently, such systems fall outside the scope of product liability systems (*see* Dou, 2025).

Second, regarding the functional positioning of AI systems for clinical decision-making, a dichotomy exists between *auxiliary* and *substitute* roles, with the former being the prevailing view. This distinction reflects corresponding optimistic and pessimistic tendencies regarding the development of such systems. Specifically, while system development is expected to enhance medical efficiency and improve healthcare standards, it also raises concerns about increased reliance on AI among medical professionals, erosion of clinical skills, and diminished accountability (*see* Li, 2020, p. 54; Li and Shi, 2019, p. 93). This analysis of functional positioning is not merely descriptive but also normative. While research acknowledges the tension between the system's high autonomy and strong decision-making capabilities versus its auxiliary role (*see* Zheng, 2023, pp. 206–207), current policies and studies consciously reinforce healthcare professionals' duty of independent review through this restrictive functional positioning. This approach aims to prevent harm from the indiscriminate application of high-risk AI. However, as the counterpart to *substitute*, the meaning of *auxiliary* remains rather ambiguous. While this functional positioning helps reinforce healthcare professionals' duty of independent review, it fails to clearly delineate the internal division of labour and interaction patterns between AI systems and medical personnel. It also falls short of articulating concrete pathways to reconcile promoting technological innovation with ensuring healthcare professionals exercise autonomous and responsible clinical judgement.

Third, the standard of care required of responsible parties is closely tied to the aforementioned issue of functional positioning. The key lies in allocating responsibility between providers and users of clinical decision-making AI systems. This involves determining the standard of conduct required of each party throughout the entire process—from development to deployment—to be considered appropriately fulfilling their responsibilities, as well as assigning specific liability to each party in the event of actual harm. Both the decision-making capability of AI systems and the healthcare professional's duty of independent review are regarded as critical elements in the clinical

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 16 Nov 2026 at 18:59:47, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/als.2026.10070

decision-making process. Therefore, the *all-or-nothing* test standard—where the absence of one element negates a finding of causation—proves inadequate for determining the factual causation attributable to either AI or the healthcare professional (*see* Hou, 2025, p. 70). Moreover, legal practice regarding causation assessment extends beyond factual causation to encompass *reasonableness judgements*. This involves further examining whether a particular factual causal relationship aligns with ordinary course of events, thereby avoiding the imposition of unduly harsh liability. This approach safeguards freedom of action and strikes a balance between risk and innovation in medical development (*see* Dou, 2024, p. 66). On the one hand, healthcare professionals face heightened expectations of responsibility. For instance, some studies advocate including the improper use of AI systems by medical personnel within the scope of medical malpractice (*see* Li and Shi, 2019, p. 95). On the other hand, the decision-making capabilities of AI systems are also factored into liability assessments. Some studies even expect technology providers to incorporate healthcare professionals' general negligence as a foreseeable risk into system development considerations (*see* Hou, 2025, p. 71; Zheng, 2023, p. 220). The rationale behind such views is to prevent healthcare professionals from becoming complacent or shirking responsibility, while simultaneously maximising the value of AI technology to improve medical accuracy and achieve reasonable risk allocation. Theoretically, imposing heightened standards of care on all involved parties could maximise risk control. However, in practice, simultaneously elevating responsibilities across all stakeholders risks creating ambiguous liability. This ambiguity may trigger a *bystander effect* where parties deflect responsibility, lacking sufficient incentive to assume accountability—ultimately undermining positive human–machine interaction and technological advancement. It is worth mentioning that there have been no publicly available court rulings in China regarding liability for damages caused by AI-assisted medical diagnosis and treatment. A 2023 review of England and Wales and a 2024 review published by the American Osteopathic Association also reported that they could find no judicial decisions turning directly on medical malpractice arising from the use of a clinical decision support AI system (*see* Jones, Thornton and Wyatt, 2023, p. 510; Shumway and Hartman, 2024). As the use of AI systems becomes more widespread, the issue of responsibility has become urgent. The scarcity of relevant reported judgements should itself be a matter of serious concern.

Fourth, regarding the liability model for compensating damages after harm occurs, existing research presents a dichotomy between "reverting to individualized liability" and "establishing independent system liability." The former approach, grounded in traditional legal frameworks, advocates that technology providers, sellers, medical institutions, and healthcare personnel bear liability for damages according to product liability and medical malpractice regulations. The latter proposes establishing an independent liability system through risk pools, compensation funds, or mandatory insurance to comprehensively address damages. It is worth noting that given the latter approach's advantage in providing clarity regarding compensation, the former is often viewed as a stopgap measure until the latter matures, aiming to gradually pave the way for specialised financial support (*see* Cai, 2023, pp. 125–126; Heine and Quintavalla, 2024). However, constructing an independent compensation system—especially with capped claim amounts—may not necessarily encourage more responsible behaviour. Such an approach focuses exclusively on the final payment rather than on damage prevention or the AI development and application process itself. This risks diverting attention from the root causes of harm and ultimately devalues responsibility.

Existing research is not entirely unaware of the unique challenges posed by such systems and the limitations of traditional liability frameworks in their application. The reliance on traditional systems is functionally oriented, seeking to leverage existing institutional designs with clear attribution of responsibility to strike a value balance between encouraging AI technological innovation, ensuring healthcare professionals'

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 16 Nov 2026 at 18:59:47, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/als.2026.10070

autonomous and responsible medical practice, and safeguarding patient rights. However, in terms of specific implementation pathways, the preceding analytical frameworks and liability allocation schemes fail to resolve these tensions. This failure arises because these liability systems and analyses perpetuate the logical framework of existing legal systems—relying on a linear causality and mechanistic worldview to achieve the clarity and certainty sought by legal institutions—and thus fail to adequately align with the operational characteristics of clinical decision-making AI systems.

Specifically, first, the development and operation of such systems have already moved beyond the assumptions of traditional product liability. Existing product liability systems presuppose that humans can consciously define a product's functional objectives and establish clear design pathways, and that harmful consequences are traceable due to well-defined causal chains. For this reason, one can retroactively determine whether producers at each stage of the production line fulfilled their due diligence obligations under the defect-based liability standard, ensuring that the respective responsibilities of relevant parties remain proportionate and foreseeable. As scholars have noted, while product liability appears to be a strict liability regime, it actually incorporates fault analysis into the determination of product defects. In other words, product defects represent the objectification of fault determination (*see* Chen, 2025, p. 376). However, the development and application of AI systems do not align with these characteristics. First, attempts to exhaustively identify liability subjects through product liability schemes are unlikely to succeed. For example, such AI systems rely on massive databases. Inherent biases in data inputs may stem from deeply ingrained societal perceptions, making it difficult to identify clear responsible parties or to address these biases solely through technical providers. Second, the emergent nature of decision outcomes and the black-box characteristics of algorithms prevent clear identification of causal links between inputs and outputs, thereby hindering comprehensive control. Yet it is precisely these traits that enable AI systems to demonstrate creativity and provide valuable insights to humans. Accordingly, the development of such technology cannot and should not be predicated on total controllability. Effective regulation of technology should not be a simple return to the logic of traditional liability systems.

Second, the functional positioning of clinical decision-making AI systems and the allocation of duty of care among responsible parties often reflect a path-dependent understanding of human–machine relations through master–servant dynamics. Those expressing vigilance and pessimism towards AI application particularly tend to presuppose a linear and modular decision-making process. Within this framework, medical professionals are assumed merely to accept decisions generated by AI and to be ultimately fully replaced by technological advancement. To ensure human autonomy, existing research counters narratives of AI domination by emphasising the finality of medical professionals' judgements. Yet this counter-narrative fails to move beyond the master–servant framework, overlooking the inevitable interpenetration of human and AI judgements in actual decision-making processes and the potential for symbiotic collaboration. Consequently, it cannot adequately reconcile technological innovation and autonomous medical responsibility.

Third, the preceding proposal, which establishes an independent fund to guarantee compensation as the blueprint for a liability system, reflects a regulatory approach prioritising compensation outcomes over practical processes. It continues to ground the liability system in a logic of retrospective causality, failing to fully consider the system's impact on future actions or to focus institutional design on advancing future technological development and refining medical practice processes. As previously demonstrated, compensation-centred approaches have inherent limitations in regulating stakeholder behaviour. Limited liability may paradoxically encourage irresponsible practices, hindering direct confrontation and resolution of issues.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 16 Nov 2026 at 18:59:47, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/als.2026.10070

3. The human–machine association as the basic decision-making unit and relational responsibility

Therefore, to establish a more effective responsibility system, we must first dismantle existing frameworks and proceed as follows. First, transcend the preceding subject–object dichotomy and the perception of human–machine relations rooted in master–servant dynamics. Second, acknowledge the inexhaustibility of responsible entities, the existence of algorithmic black boxes, and the failure of linear causal logic, thereby ceasing to ground responsibility systems upon determinism. While achieving algorithmic transparency remains a worthy goal, the practical application of clinical decision-making AI systems and the pursuit of responsibility should not depend on such transparency. Consequently, it is essential for legal systems to develop the capacity to promote responsible actions and ensure effective accountability without requiring humans to see through AI systems. Third, we must recognise the interpenetrating nature of human–machine collaborative decision-making. Such processes cannot be reduced to simple aggregates of individual decision-making practices, thereby making atomised attribution inadequate. Opportunities for symbiotic human–machine cooperation should be sought, fostering more efficient and responsible actions through mutual empowerment rather than allowing one to undermine the other. Fourth, while addressing final compensation, greater emphasis should be placed on the dynamic process itself and the establishment of future-oriented responsibility frameworks. Consequently, the legitimacy of responsibility rests less on the establishment of factual causation or the blameworthiness of the agent, and more on the positive construction of social relations and the normative significance of reconciling conflicting interests and values.

3.1. The human–machine association as the fundamental decision-making unit

This article argues that the nature and functional positioning of clinical decision-making AI systems are not determined by fixed physical attributes, but rather are constituted through human–machine interaction (*see* Ji, 2020, pp. 60–61). Analysis of such systems should focus not on their *ontological reality*, but on how they appear to humans during interaction, and on humans' corresponding perceptions, experiences, and responses (*see* Wang, 2024, p. 20). In the decision-making process, system outputs and clinicians' judgements become intertwined, and the final decision cannot be reduced to either party's individual decision or a simple aggregate of both. Therefore, the *human–machine association* should be regarded as the fundamental unit of clinical decision-making behaviour (*see* Beckers and Teubner, 2023).

In practice, the patterns and degrees of human–machine decision integration vary. Medical practitioners' diagnostic workflows typically encompass rule-based and heuristic processes. Rule-based processes follow explicit clinical guidelines, offering standardisation and reproducibility, and are often applied to conditions with established research. Heuristic processes, however, rely primarily on patterned cognitive frameworks—even intuition—developed through prolonged learning and practice, resulting in a black-box decision-making process. In this process, clinicians rapidly form preliminary hypotheses by associating similar diseases, recalling past cases or events, or anchoring specific information. Heuristic approaches are often applied to complex conditions (*see* Terry, 2019, pp. 139–140), holding significant implications for medical advancement. However, conclusions derived through this method still require reflective validation. Notably, these two diagnostic approaches align with AI systems based on rule-based algorithms and deep learning. Generally, simpler conditions involve more well-established medical knowledge and clearer diagnostic standards. In such cases,

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 16 Nov 2026 at 18:59:47, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/als.2026.10070

AI systems should possess the capability to perform a stable, validated task within an explicit use envelope, and independently execute complete decision-making processes, while clinicians' subsequent judgements can often build upon the diagnostic pathways provided by the AI system. Here, the relationship between AI and clinicians resembles a linear process of *AI autonomous decision-making + clinician review*. The system exhibits strong autonomy, with minimal interpenetration between human and machine judgements, closely approximating an agency model (where AI independently replaces human decision-making). Yet, since clinicians remain involved, this does not constitute true agency. In comparison, the diagnosis of complex conditions relies more heavily on deeper interaction and communication between AI and medical professionals, with final conclusions emerging from mutually stimulating cognitive processes. Current technological trends also validate this approach. For instance, RDguru, a conversational AI diagnostic model, narrows differential diagnoses through multi-round dialogue mechanisms, aiding rare disease diagnosis (*see* Yang et al., 2025). Similarly, the KAMAC system achieves breakthroughs in complex clinical scenarios by emulating multidisciplinary expert team collaboration in real medical decision-making. This enables large language model agents to initiate knowledge-driven discussions with one or more expert agents, dynamically assemble and expand expert teams based on evolving clinical contexts, review and update decisions, and ultimately reach conclusions (*see* Wu et al., 2025). Similarly, the core concept behind Microsoft's newly developed MAI-DxO clinical decision support system draws inspiration from the collaborative patterns observed in real medical teams (*see* AIGC Open Community, 2025). This demonstrates that communication and interaction models play a pivotal role in advancing technological development and enhancing clinical decision-making capabilities. As AI technology evolves and is applied to increasingly complex problems, its interactive characteristics will inevitably become increasingly prominent.

In summary, the relationship between clinical decision-making AI systems and their users forms a spectrum ranging from *partnerships approaching agency* to *deeply interactive partnerships*. The deeper the interaction, the less distinct the division of labour between humans and machines becomes, and the less reducible the final decision.[1] It is precisely through this focus on the unique characteristics of human–machine interactive decision-making that we can more appropriately understand the *auxiliary* functional positioning of clinical decision-making AI systems. Auxiliary function does not establish a master–servant relationship between humans and machines, nor does it mean that medical professionals' treatment plans and value standards override AI judgements. Auxiliary function signifies, on one hand, recognition of AI's autonomous decision-making capabilities, and, on the other hand, the need for medical professionals to make the final judgement on the understandability and acceptability of treatment plans. This process of understanding and acceptance is not disconnected from the AI system's decision-making. Through this interaction process, medical professionals' information and knowledge have been expanded, and their diagnostic decision-making assumptions have been calibrated. Therefore, the auxiliary position of AI ensures the presence of medical professionals while avoiding rigid adherence to their final decision-making authority. Instead, it calls for context-specific configurations of responsibility, calibrated to the characteristics of human–machine interaction.

[1] The criteria for distinguishing between agency relationships and association relationships are primarily reflected in the following three aspects: first, whether fault can be clearly attributed to a decision made by one party; second, the degree of collaboration between humans and algorithms; third, whether the interaction pattern involves bilateral cooperation or unilateral directives. Of course, in real-world practice, grey areas always exist, making it impossible to draw a clear-cut distinction. *See* Beckers and Teubner (2021), p. 154.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 16 Nov 2026 at 18:59:47, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/als.2026.10070

3.2. Establishing a relational responsibility system

Furthermore, treating the human–machine association as the basic decision-making unit does not mean that the association as a whole should bear responsibility. Otherwise, it risks repeating the pitfalls of schemes like "establishing independent compensation funds," which serve merely to satisfy financial compensation requirements without improving practice. Simultaneously, the framework must avoid atomised responsibility arrangements that fragment the actions of various actors and their causal contributions to final outcomes.

Relational responsibility offers a third approach to overcome these challenges. This analytical approach emphasises that while we commonly express responsibility as "being responsible for something," its essence lies in "being responsible to others" (*see* Painter-Morland, 2006, p. 93). A responsible actor is not an atomised entity but one embedded within relational contexts. Social facts emerge through the interactions of various agents and cannot be reduced to a simple aggregation of the causal forces attributed to each party. The relational responsibility system has two key implications: first, it determines the content of responsibility within the specific context of relational interaction, thereby transcending a mechanistic understanding of responsibility attribution. It treats the content of responsibility as a social construction grounded in particular values, rather than a natural fact. Second, it establishes relational regulation as the value principle of the responsibility system. The system therefore operates both retrospectively and prospectively. It is concerned not only with tracing causality, distinguishing right from wrong, and assigning blame, but also with improving the interactive process itself. Even compensation for harmful consequences is viewed as a component of relationship regulation. People do not merely seek redress for specific outcomes through the responsibility system. They focus on the expectations, understandings, and attitudes towards the other party reflected in the specific acts of assuming responsibility. Both the practice itself and the compensation for damages should facilitate the possibility of sustained interaction and cooperation among parties, rather than exacerbating confrontation and shifting blame. This shifts the value principle from *first-order morality* to *second-order morality*, and the justification for assuming responsibility from outcome-based standards of right and wrong to process-based standards that assess whether actions create space for future interaction. As social networks become increasingly complex, individual actors exercise less control over the specific outcomes generated by interactive processes. Rather than persisting in the pursuit of linear causality, we should focus on the relational interaction process itself, as this is an effective strategy for addressing the emergent nature of outcomes and the ambiguity of fault. This approach helps stimulate actors' potential to respond to risks through interactive practice (*see* Gergen, 2011).

In the specific application scenario of clinical decision-making AI systems, the emphasis on relational responsibility implies acknowledging that specific outcomes result from the interplay between the technology provider's development and production, the system's operational mechanisms, and the usage practices of healthcare institutions and clinicians. Nevertheless, it still holds the technology provider and users accountable as concrete responsible parties, requiring them to exercise their agency. In joint decision-making scenarios, the causal contributions of each party may not be clearly delineated. However, society cannot function through mutual shirking of responsibility and indifference. Such an approach neither addresses the pressing reality of scarce healthcare resources nor enhances medical standards through beneficial human–machine interaction, thereby safeguarding patient interests. Therefore, the central task is defining the desired mode of interaction and advancing its realisation through a framework of responsibility. Guided by relational responsibility theory, responsibility is oriented towards the relational patterns people expect, with the institutional goal of regulating how and with what attitude

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 16 Nov 2026 at 18:59:47, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/als.2026.10070

relational actors treat each other. The specific content of responsibility must be determined in light of particular human–machine interaction relationships and contexts, rather than being an explicit, static, object-centred standard, as exemplified by product defects. Moreover, it is precisely people's focus on the specific meanings and attitudes underlying relationships and actions that indicates the insufficiency of current AI systems, which lack specific volition, attitudes, and emotional response, and by extension, the insufficiency of the human–machine association itself to bear the responsibility people expect. Only by directly addressing active agents such as technology providers and users can the system achieve regulation of the relationship itself.

This conception should also be distinguished from two established relational models in law. Fiduciary law begins from entrusted discretionary power, vulnerability, and confidence, and responds principally through duties of loyalty and remedies for betrayal of the beneficiary's trust. Its bilateral and asymmetrical structure illuminates the physician–patient relationship, but it does not by itself organise the polycentric interaction among developers, institutions, practitioners and patients across an AI lifecycle (*see* Unver, 2023, pp. 198–200). Relational contract theory, by contrast, correctly treats exchange as embedded in an ongoing social relationship sustained by cooperation, adjustment, and norms that exceed a discrete promise (*see* Feinman, 2000). Yet its primary field remains the continuing exchange relation between contracting parties. The relational responsibility proposed here borrows the fiduciary concern with vulnerability and the relational-contract concern with continuing cooperation, but extends beyond both: it includes non-contracting patients and institutional actors, makes the human–machine association the analytical unit, and evaluates responsibility by whether lifecycle practices construct, maintain, and repair relationships necessary for clinical decision-making. Its distinct contribution is therefore not the claim that every AI relationship is fiduciary or contractual, but the use of a broader, future-oriented responsibility principle to coordinate the legal duties that arise from several different sources.

In summary, the preceding analysis treats the human–machine association as the fundamental unit of clinical decision-making, adopting a spectrum-based understanding of how AI systems for clinical decision-making operate. It then embeds decision stakeholders within their interactive networks and imposes relational responsibility. This two-step approach offers three advantages and provides direction for constructing concrete responsibility arrangements: first, treating the decision association as the basic unit helps prevent responsibility shifting. Neither the inherent complexity and black-box nature of the technology nor the healthcare provider's duty of independent review can justify mutual shifting of responsibility between technology providers and users. This approach avoids perpetuating the attribution logic of traditional product liability systems while addressing a core concern in existing research: including all decision-involved parties within the scope of responsibility. Second, a spectrum-based account of the depth of human–machine decision integration refines the interpretation of *auxiliary role*, transcending the master–servant relationship analytical framework. This avoids framing technological advancement and medical autonomy as opposing forces in a zero-sum dynamic, instead promoting their mutual empowerment and coordination. Third, the specific interaction model adopted within the decision-making association relates not only to the development of AI systems and their practical implementation tailored to clinical needs, but also to society's overall expectations for human–machine interaction and the judgements and decisions made by clinicians in specific clinical practices. Therefore, the design of a responsibility system for human–machine association extends beyond compensation for harmful consequences and should encompass the practical process itself. This also represents a breakthrough from both technocentrism and anthropocentrism, thereby neither reducing real-world dilemmas to mere technical advancement nor viewing technology solely as a tool manipulated by humans. Rather than focusing on the

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 16 Nov 2026 at 18:59:47, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/als.2026.10070

static attributes of AI systems themselves, this framework prioritises the interaction between technology and society, thereby constructing an appropriate responsibility system grounded in specific interactive contexts and genuine human needs (*see* Li, 2016, p. 44).

4. Constructing a responsibility framework for AI applications in clinical decision-making from the perspective of trust relationships

Building upon the two-step responsibility framework, the construction of responsibility systems comprises two key components. First, how the human–machine association—and the technology providers and users constituting the relational subjects within this association—should assume external responsibility? Second, what mutual responsibility expectations should exist within the human–machine association to guide their interactions? The following discussion shows that these institutional requirements call for a focus on trust relationships. Consequently, the two preceding issues become questions of how to build physician–patient and algorithmic trust within the context of AI-assisted clinical decision-making. This trust-based approach incorporates real-world uncertainties into institutional design considerations, treating *relationships* themselves as fundamental regulatory units, and focusing on concrete practices throughout the entire AI lifecycle from technology development to application.

4.1. From atomic responsibility allocation to trust relationship building

As previously noted, the traditional principles, which emphasise clear linear causation and atomised responsibility, are inadequate for regulating human–machine interaction, and may encourage responsibility shifting. Overcoming this bottleneck requires attention to the reality of human–machine interaction and a shift in analytical perspective from individuality to relationality. From a relational perspective, the arrangement of duty of care can be reinterpreted as determining the scope of mutual trust among relational participants. Specifically, the duty of care imposed on one responsible party should correspond to the degree to which it can trust other responsible parties, that is, the extent to which it need not exert additional vigilance. It is precisely upon this foundation of trust that all parties act, and collectively form the final decision.

It is worth clarifying the meaning of *trust* to demonstrate the value of *trust relationships* perspective. As Simmel noted, trust is one of the most important synthetic forces in society (*see* Simmel, 2002, p. 251), which is a mixture of *knowing* and *not knowing*, meaning that even when the parties are aware of the lack of complete control over trustees' actions, they are still willing to assume that trustees will act in a manner beneficial to them (*see* Dai and Wang, 2025, p. 40). Thus, a trust relationship implies accepting the existence of information asymmetry, opacity, and uncertainty while maintaining rationality as the basis for action under these conditions. Trust does rest on some degree of assessment, but it does not depend entirely on epistemological factors; that is, it is not founded on complete controllability or explicability. A *leap of faith* is inevitably required between acquiring knowability and explicability and taking cooperative action. With trust relationships as its normative goal, legitimacy for actions undertaken by all parties derives not from explicit causality or individual fault. Rather, it lies in the capacity of such actions to promote a constructive trust relationship oriented towards the practical process itself and future development. The construction of responsibility systems does not presuppose transparency and controllability and is not aimed solely at compensating for harmful consequences. Rather, it seeks to advance the appropriate interactive process itself, enabling technology providers, users and patients to engage in diagnostic and therapeutic

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 16 Nov 2026 at 18:59:47, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/als.2026.10070

activities with mutual trust despite the realities of algorithmic black boxes and medical risks. Technological maturity, the establishment of clear standards, and the assumption of liability for damages are essential elements and the rational basis for building trust. Therefore, the preceding shift in the legitimacy basis of responsibility does not displace requirements of traditional liability systems. However, considering the practical context of AI applications in clinical decision-making, responsibility frameworks should not merely regulate actions with certainty. They must also ease value tensions within conditions of uncertainty, thereby creating space for practice to explore the optimal blend of *knowing* and *not knowing*.

By contrast, the weakness of existing frameworks lies precisely in their failure to ground the system in this hybrid of *knowing* and *not knowing*. When prior research treats algorithmic black boxes as grounds for exempting medical providers from liability, it accordingly pursues the explainability of AI systems. Such explanations primarily manifest in three modes: first, contrastive explanation, justifying why the system produced a specific outcome rather than others, and demonstrating why the generated result is more reasonable; second, functional explanations, showing what output the system derives from a given input to perform a specific function; third, transparency explanations, disclosing the system's internal structure and operational mechanisms. These explanatory approaches reflect efforts to address algorithmic black boxes. For example, functional explanations can be provided even without understanding the decision logic. However, these approaches all focus on assessing the correctness of decision outcomes. They assume users will become willing to use the system based on such outcome judgements. The problem lies in users' limited ability to assess outcome accuracy. Such post-hoc explanations are also easily manipulated as a rhetorical tool. For instance, comparison items can be manipulated to create the illusion of credible results. In contrast, the advantage of building responsibility systems around trust relationships has two dimensions. First, the trust relationship framework aligns with the current reality of diverse, intertwined actors and algorithmic opacity, advancing technological practice on this basis. It does not advocate eliminating the inherent knowledge and information gaps between system users and developers, or between medical professionals and patients. Instead, it seeks to guide cooperative directions within these disparities, creating space for technological application. Second, establishing trust relationships concerns not only the rationality of decision outcomes and liability after harm occurs, but also the interactive process from system production to use. This is particularly significant for addressing the development and practical challenges of AI technology (*see* Bao and Zeng, 2024).

As previously noted, the responsibility framework here encompasses both physician–patient trust and algorithmic trust in clinical decision-making AI systems. The former involves external responsibility towards patients, fostering collaboration to advance treatment plans. The latter constitutes internal responsibility arrangements between technology providers and users through human–machine interaction, enabling the human–machine alliance to deliver effective clinical decisions. This distinction does not imply separate operational processes for the two trust relationships. Rather, they interact and reinforce one another in practice. For example, only when trust exists between physicians and patients can physicians access more effective patient information and specific treatment needs, thereby enhancing the responsiveness of human–machine interactive decisions and consequently strengthening algorithmic trust. Simultaneously, only by establishing algorithmic trust can healthcare professionals provide clearer and more persuasive explanations of clinical decisions to patients, which in turn helps deepen physician–patient trust.

In practice, diverse foundations for trust have emerged. Specifically, trust may rest on the following bases, ranging from personalised experiences to macro-level social structures, and from particular to universal contexts: (1) based on understanding of

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 16 Nov 2026 at 18:59:47, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/als.2026.10070

others and calculation of interests, such as determining through assessing the trustees' motivation to honour commitments and safeguard the parties' interests; (2) based on specific interactive relationships, such as gradually entrusting one's interests to another through long-term engagement; (3) based on others' group identity, social roles, and position within social networks, such as trust arising from shared ethnicity or professional affiliation; (4) based on institutional arrangements, which provide the foundation for individuals to advance relationships and make decisions when social complexity makes full interaction difficult yet decision-making is necessary; (5) based on behavioural norms, such as the belief in a universally accepted code of mutual trust within society. Importantly, these forms of trust are not grounded solely in rational analysis but are also influenced by emotions. For example, trust developed through long-term interactions is both a rational judgement derived from inductive knowledge and a leap of faith ultimately realised through emotional impetus. The trustee does not decide to honour commitments based solely on rational calculation. If trust is betrayed, the trustees feel they have violated mutual expectations for the relationship, thereby damaging their self-identity. Moreover, specific trust and general trust are not interchangeable. Institutional arrangements, for instance, cannot replace the specific trust formed through concrete interactions. Some scholars even contend that relationships built on institutional arrangements—relying on the clarity of rules and predictable consequences—differ from trust relationships that exist in the grey area between *knowing* and *not knowing*. Therefore, institutions alone cannot serve as the driving mechanism for trust relationships.

For physician–patient and algorithmic trust, the respective foundations can be found in the preceding account. Trust between physicians and patients may rest on: (1) patients' understanding of physicians' professional competence; (2) long-term interactions between physicians and patients during the diagnosis and treatment process; (3) physicians' tendency towards responsible medical practice due to their professional identity, manifesting both as the constraining effect of professional ethics as role norms and as a reflective self-identity possessed by physicians; (4) institutional norms governing the physician–patient relationship and a prevailing social climate of trust between physicians and patients. Algorithmic trust, meanwhile, may rest on the following factors: (1) people's understanding of artificial intelligence systems, specifically the system's level of explainability; (2) the accumulation of long-term human–machine interaction experiences; (3) Relevant institutional regulations and a socially prevalent environment of trust in algorithms. Overall, algorithmic trust primarily rests on functional attributes and inductive judgements formed through prolonged interaction, whereas physician–patient trust involves emotional factors beyond rationality and hinges on holistic assessments of the practitioner's benevolent intent. This emphasis on emotion and intent is crucial in motivating individuals to engage in risk-taking activities.

Because of the distinctive features of AI applications in clinical decision-making, three points warrant particular attention: first, we should acknowledge the reality of algorithmic black boxes. While high accuracy in long-term human–machine interactions can enhance trust, such trust is fragile and prone to collapse when biases emerge. Moreover, the critical nature of medical issues means they cannot be evaluated solely by the probability of algorithmic accuracy. Therefore, the building of algorithmic trust must transcend a model that relies solely on evaluating the technology itself. Second, regarding physician–patient trust, patients place trust in physicians not only based on specific interactions but also on physicians' inherent benevolent intent associated with the medical profession. Accordingly, healthcare professionals derive a sense of self-identity from their role as physicians, engage in self-reflection and develop emotional attitudes based on whether they have earned and maintained this trust. The inherent emotions and attitudes reflect the distinctiveness of human-to-human interaction compared to human–machine interaction. In the absence of knowledge certainty, this is particularly significant for

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 16 Nov 2026 at 18:59:47, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/als.2026.10070

patients' willingness to trust and cooperate with medical care. Therefore, in arranging responsibilities for medical institutions and practitioners, it remains essential to reinforce this unique foundation of physician–patient trust, rather than advocating for the withdrawal of physicians solely on the grounds of technological advancement. Moreover, the reflexivity and emotional elements inherent in interpersonal trust can compensate for the limitations of algorithmic trust, which is primarily based on cognition. This, in turn, supports technological development by enabling healthcare professionals to mediate and foster patient trust in the application of artificial intelligence systems within medical processes. Third, the establishment of a responsibility system as a macro-level institutional arrangement should not merely aim to regulate the behaviour of responsible parties from the top down, providing rule-based certainty for legal consequences. Rather, it should facilitate spontaneous interaction among all parties, fostering flexible trust relationships responsive to real-world contexts. This aligns with the agency inherent in the relational subjects discussed earlier and the emergent nature of human–machine collaborative decision-making.

4.2. External responsibility to foster physician–patient trust

With respect to external responsibility, it is particularly crucial to leverage AI systems to support trust-building between physicians and patients, and avoid the emergence of new forms of medical paternalism due to technological integration (*see* Huang and Tan, 2024, pp. 51–52). This responsibility system has two dimensions: first, the dialogic responsibility of medical providers during the diagnostic and treatment decision-making process; second, the primary responsibility of medical providers as the accountability hub when harm occurs.

First is the issue of responsibility allocation in the clinical decision-making process. It can be observed that the introduction of relational responsibility has updated and deepened the meaning of *informed consent*, which gives practical effect to Article 1219 of China's Civil Code.[2] Specifically, prior to making a clinical decision, healthcare institutions should explicitly explain to patients the specific interaction model between AI and medical personnel during AI-assisted diagnosis and treatment, as well as the associated benefits and risks. To be concrete, the validity of *informed consent* under AI opacity cannot depend on explaining source code and every parameter. Such a requirement would overwhelm patients without necessarily improving autonomy. Nor, however, should the difficulty of technical explanation justify silence. A relational approach calls for layered, continuing communication. At the baseline, the physician should disclose that AI materially participates in the recommendation, the function it performs, the degree of human review, known material limitations or population mismatch, reasonably available non-AI alternatives, and the consequences of accepting or declining its use. Communication intensity should increase where the system is unusually opaque and novel, where the clinician cannot independently verify a material output, or where the patient asks about AI involvement. This empowers patients to autonomously decide whether to incorporate AI-assisted diagnosis and treatment. During the treatment process, if a patient consents to using a clinical decision-making AI system, the system's deep learning capabilities and

[2] Article 1219 of the Civil Code of the People's Republic of China: Medical personnel shall explain the patient's condition and medical measures during diagnosis and treatment activities (2020). When surgery, special examinations, or special treatments are required, medical personnel shall promptly explain specific medical risks, alternative treatment options, and other relevant circumstances to the patient and obtain their explicit consent. Where it is impossible or inappropriate to explain to the patient, medical personnel shall explain to the patient's close relatives and obtain their explicit consent. Where medical personnel fail to fulfil the obligations set forth in the preceding paragraph and cause harm to the patient, the medical institution shall bear liability for compensation.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 16 Nov 2026 at 18:59:47, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/als.2026.10070

algorithmic black-box nature still do not constitute grounds for exempting medical personnel from liability. Medical professionals should adhere to the principles of relational responsibility, engaging in dialogue with patients about treatment decisions. This involves understanding the patient's expectations regarding treatment goals and concerns about the proposed plan, and providing further clarification based on this understanding.[3] Implementation of dialogic responsibility offers the following advantages: first, dialogic responsibility facilitates healthcare providers' reflection on their own decision-making processes, which prevents them from blindly trusting AI conclusions based solely on historical accuracy, thereby reducing bias. Second, strengthening doctor–patient dialogue aligns with the practical demands arising from the increasing prevalence of chronic diseases. In such cases, curing the disease is not—and often cannot be—the ultimate goal of treatment. Patient expectations for treatment are diversifying—some may prioritise pain reduction over cure, while others seek to maintain functional abilities like eating or mobility. Treatment plans cannot be predetermined without patient input. Rather, they must emerge from collaborative, iterative dialogue to identify and reflect patients' value priorities in clinical decisions. Such dialogue also facilitates feedback to AI systems, preventing them from generating treatment plans that contradict patients' actual wishes. Third, direct doctor–patient dialogue strengthens trust between them. While AI systems may exhibit high decision accuracy in statistical and probabilistic terms, in medical contexts, each patient's rights are paramount. Therefore, trust cannot be built solely through probability. As previously noted, the mechanisms for building interpersonal trust and algorithmic trust differ. Concrete physician–patient communication can stimulate human reflexivity and emotional elements. This is crucial for trust-building and aligns with the emerging medical ethics concept of *narrative medicine*. Narrative medicine aims to enhance patient engagement in their illness, providing opportunities to articulate the broader social contexts underlying their treatment needs. This encourages physicians to listen to and understand patients' stories, thereby transcending the limitations of the biomedical model, which focuses solely on combating disease while neglecting the patient as a whole (*see* Jiang and Lin, 2024). Overall, the introduction of dialogic responsibility implies that informed consent is not a one-way information dump but an iterative relationship-building practice. The physician–patient relationship can no longer be conceived in unidirectional term, and thus cannot be reduced to an either/or choice between physician-based and patient-based disclosure standards. Instead, shared decision-making can be achieved through physician–patient dialogue.

One possible objection to this dialogic responsibility is that in traditional medical practice, healthcare providers are not obligated to disclose to patients the literature or case studies referenced in their decision-making, nor to reveal their reasoning process. Given this, why should healthcare providers be obligated to explain their specific interaction patterns with decision-making AI systems to patients (*see* Cohen, 2020, p. 1442)? This article argues that the process of a physician conducting literature searches based on their preliminary analysis of clinical symptoms differs from their direct interaction with AI systems. The former involves a thinking framework dominated by the clinician, while the latter represents an interaction between two thinking frameworks, making clinicians more susceptible to anchoring effects. This influence is often difficult to avoid and can be mitigated only through the preceding reflective dialogue. Therefore, although both the healthcare professionals' literature search and the application of an AI system serve an auxiliary role, the nature and operation of assistance remain fundamentally different. In scenarios where AI systems are used to support decision-

[3] In fact, dialogic responsibility constitutes the core element of the relational responsibility system. *See* McNamee and Gergen (1999), pp. 5–6.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 16 Nov 2026 at 18:59:47, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/als.2026.10070

making, this dialogic responsibility plays a crucial role in strengthening both the rational and emotional foundations of trust between physician and patient.

The second issue concerns liability allocation when medical decisions result in harm. As previously discussed, the human–machine association is the fundamental unit for clinical decision-making. However, its liability is limited to monetary compensation. To treat the human–machine association alone as the liable party is to reduce the responsibility system to a punitive instrument, obscuring its more fundamental role as a vehicle for cultivating relational trust and fostering constructive engagement among the parties involved. Therefore, this article advocates a two-step approach to implementing liability for harm: first, theoretically establish the human–machine association as the fundamental unit for decision-making liability. In this case, patients need not prove the factual causation and degree of fault of individual technology providers, sellers, medical institutions, and healthcare personnel to seek separate compensation. Instead, they only need to demonstrate a causal link between the final treatment decision made by the human–machine association and the resulting harm. They must also demonstrate that this final decision was negligent, considering the prevailing medical standards and the patient's specific condition as reflected in the history of physician–patient interactions. Second, regarding the specific entity liable for compensation, the medical institution serves as the pivotal hub within this system, centrally assuming external liability. Subsequently, the medical institution may internally allocate responsibility among other liable parties based on specific human–machine interaction. The internal allocation of liability will be detailed later and is not elaborated upon here.

The rationale for this two-step approach rests on two considerations: first, treating the human–machine association as the theoretically liable entity aligns with the emergent nature of interactive decision-making, which cannot be reduced to the individual actions of any single participant. Moreover, even if medical practitioners assume dialogic responsibility during the diagnostic decision-making process by explaining their interaction patterns with AI systems to patients, patients still struggle to comprehend the actual internal mechanisms of human–machine interaction. Requiring patients to pursue compensation based on relational characteristics among various entities is impractical. Designating the human–machine association as the theoretically liable entity substantially reduces the burden of proof for patients. Second, designating the medical institution as the primary recourse entity not only facilitates patient compensation but also serves as an institutional solution for building trust between patients and healthcare providers, as well as trust in algorithms. This creates a consistent framework for *physician presence*, encompassing prudent reflection on treatment decisions during active dialogue and taking the lead in assuming responsibility after treatment-related harm. This approach adheres to the *human-in-the-loop* model of human–machine interaction, ensuring that while AI systems permeate human decision-making, human judgements retain primacy in specific decisions. The legitimacy of healthcare institutions assuming centralised liability in the first instance stems neither from institutional fault nor punitive intent, but rather from fostering trust-building through this system and empowering algorithmic trust through physician–patient trust. This mechanism is also supported by current legislation. According to Article 1223 of China's Civil Code, when a patient suffers harm due to a defective medical device, the patient may first seek compensation from the medical institution. After paying compensation, the medical institution may then seek reimbursement from the responsible manufacturer or other liable parties. Admittedly, this provision of the Civil Code is grounded in a clear dichotomy between medical device defects and medical malpractice. Yet even in cases where fault can be distinctly categorised, the law still supports patients requesting medical institutions to bear responsibility first. By extension, imposing such liability requirements on medical institutions in interactive decision-making scenarios is justified, reflecting a consistent

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 16 Nov 2026 at 18:59:47, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/als.2026.10070

consideration for safeguarding patient rights and building trust between physicians and patients.

Naturally, the preceding mechanism does not preclude medical institutions from establishing dedicated funds for compensation. In the longer term, a specialised compensation fund or pooled coverage may address diffuse, latent, or scientifically indeterminate injuries. However, it should be clarified that although insurance and funds are valuable simplification and loss-spreading devices, they remain relationally legitimate only when designed to return information and incentives to the actors capable of improving future interaction, rather than purchasing an exit from responsibility. Payment should trigger, rather than erase further investigation and internal accountability.

In summary, the medical provider's fulfilment of dialogic responsibility and its primary assumption of liability for damages encourage careful review of treatment decisions. This approach also enhances the acceptance of medical technologies through strengthened trust between physicians and patients, thereby balancing the relationship among technological innovation, medical autonomy, and patient rights.

4.3. Internal responsibility arrangements to foster algorithm trust

A further question is how AI systems and their providers should interact with healthcare institutions and practitioners as users to foster algorithmic trust and achieve mutual empowerment between technological innovation and medical practice? In scenarios involving algorithmic black boxes, algorithmic trust cannot be built solely on the transparency and interpretability of AI systems. Instead, it requires focusing on the interaction process itself between technology providers and users. Accordingly, the internal allocation of specific liability for damages resulting from decisions must align with both parties' expectations regarding the scope of mutual trust.

As previously noted, the human–machine association is not monolithic. Its modes form a spectrum ranging from near-agency relationships to deep interactive partnerships, dynamically evolving with technological advancement. This inherently implies varying degrees of trustworthiness. Algorithmic trust within the team should therefore be understood as calibrated reliance by clinicians, not as general confidence in technology. Regarding the interactive process itself, the design intent of technology providers, system capabilities and risks, alongside practitioners' operational needs and concerns, can only be mutually clarified and aligned through dialogue. Therefore, equal emphasis should be placed on the dialogic responsibility between technology providers and users. Specifically, technology providers should articulate the underlying value norms, functionalities, and risks of the technology, the efforts made during development to address challenges, and how the technology's clinical decision-making aligns with or extends beyond healthcare practitioners' decision-making reasoning. This provides guidance on how practitioners should apply and interact with the system. Healthcare professionals, as users, should articulate their specific clinical practices for the target condition, particularly the challenges and needs encountered during practice. This enables verification of how well the technology provider's solution aligns with actual requirements. When individual clinicians decide to adopt or override the AI recommendations, the decision should be well-reasoned and logged. These logs, however, must be used for further learning rather than for automatic punishment that encourages concealment. This dialogue process must be iterative throughout subsequent practical use to drive technological advancement and evolve human–machine interaction models. For instance, if discrepancies exist between the clarity of medical knowledge and the accuracy of decision-making AI systems (e.g., AI failing to accurately diagnose conditions with well-defined clinical standards), designers should provide explanations. This helps medical practitioners understand obstacles in translating medical knowledge into algorithms while enabling technology providers to

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 16 Nov 2026 at 18:59:47, subject to the Cambridge Core terms of use , available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/als.2026.10070

grasp clinical practice needs and clarify trust levels in system decisions during human–machine interactions. This dialogic responsibility fosters shared objectives between technology providers and users, engaging users and conveying providers' concern for their needs. As previously noted, this constitutes a critical element in building trust.

It follows that interface design is also part of the dialogic responsibility arrangement: the interface should disclose the system's validated task, confidence and data limits, present clinically meaningful alternatives or counterfactuals, permit efficient correction and override, and record the reasons for adoption, rejection, and escalation. Explanations that are technically rich but poorly fitted to workflow may increase cognitive burden or become superficial badges of competence rather than triggers for reflection (*see* Naiseh et al., 2023). The logic of human collaboration with coding agents provides a useful but limited analogy. Reliable teams do not equate the human's final click with meaningful control. Instead, they decompose work into inspectable intermediate outputs, preserve version histories, permit rollback, and require heightened review before consequential deployment. Clinical AI systems should similarly expose intermediate reasoning artefacts appropriate to the task, make overrides reversible where possible, and place explicit review gates before high-risk recommendations become orders. The analogy has limits because clinical decisions concern bodily integrity, time pressure, and patient values rather than software artefacts. Precisely for that reason, the human approval point must represent substantive, supported review rather than a formal transfer of blame to the physician.

The dialogic responsibility differs from the warning duty under traditional product liability. The warning duty centres on the product itself, relying on defined design and manufacturing standards alongside foreseeable risks. It manifests as a non-interactive, one-way statement aimed at ensuring users operate within the product's intended functionality. Furthermore, proper fulfilment of the warning duty can constitute grounds for reducing or eliminating the manufacturer's liability. The dialogic responsibility, however, does not necessarily presuppose foreseeable risks or fixed usage guidelines. The appropriate manner of human–machine interaction is clarified and continuously adjusted through dialogue between technology providers and users, based on users' needs and technical feasibility. This implies that vague usage instructions provided by technology providers, such as "system-generated decisions are for reference only," cannot be considered adequate fulfilment of dialogic responsibility, and cannot negate specific commitments about validated performance or institutional directions that encourage reliance.[4] Furthermore, the dialogic responsibility differs from the interpretability of AI systems. The latter focuses on knowledge-based explanations of technology, building algorithmic trust by demonstrating the accuracy of decision outcomes. The former, however, centres on the proactive, responsible attitude manifested in the dialogue process itself and its corresponding role in constructing trust relationships. Here, trust stems not merely from users' understanding of technology but from the collaborative, interactive relationship between technology providers and users. It is worth noting that the specific consensus reached regarding the level of trustworthiness can be formalised in the purchase and usage agreements signed between technology providers and users. Given that the explicit definition of trustworthiness levels directly impacts liability arrangements in the event of subsequent harm, both technology providers and users will have sufficient incentive to practice this dialogic responsibility. Overall, by reshaping organisational conditions, these measures implement relational responsibility with a focus on the interaction process itself, rather than merely waiting to identify a single blameworthy actor after harm.

Furthermore, post-damage liability arrangements should correspondingly revolve around the calibrated level of trust. It must be reiterated that in scenarios where the

[4] In discussions with providers of such AI systems, the author has observed that they often employ such vague usage instructions to mitigate liability risks.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 16 Nov 2026 at 18:59:47, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/als.2026.10070

human–machine association serves as the fundamental decision-making unit, such liability arrangements are not based on the explicit attribution of fault or the blameworthiness of a specific party in causing harm. A party is not responsible merely because it appears somewhere in the causal chain, nor exempt merely because another party retained the formal "final decision." Instead, the party that breaches the previously established trust baseline, that is, that fails to perform the dialogue, monitoring, documentation, updating, or reflective verification reasonably expected of it, should bear responsibility for achieving relationship adjustment. Generally, if an AI system employs rule-based algorithms to mimic healthcare professionals' rule-based diagnostic processes, it implies the system possesses clear diagnostic criteria. Compared to humans susceptible to irrelevant factors and cognitive biases, such a system should inherently offer more precise diagnostic capabilities, thereby providing a stronger rational foundation for algorithmic trust. In such cases, medical professionals have greater justification for freeing up attention. Clinicians should principally check eligibility, data quality, and exceptional contraindications. Reliance within the validated envelope is more justified, so an unexplained failure within that envelope weighs more heavily in internal recourse against the AI system provider. Conversely, if the system aligns with medical professionals' heuristic diagnostic approaches through deep learning modes, the system's diagnostic decisions may either exacerbate the anchoring effect experienced by medical professionals or compensate for their oversights, offering valuable insights into treatment directions. In these instances, medical professionals should engage in more thorough deliberation and reflection on the diagnostic outcomes. Evaluation should focus on the quality of iterative questioning, independent clinical synthesis, preservation of the interaction record, disclosure of model and data limits, and the use of feedback from observed failures to inform system updating. In sum, as warranted reliance and machine autonomy increase, provider commitments and traceability should intensify. As uncertainty and interaction deepen, clinicians' reflective and dialogic responsibilities should intensify. This mechanism helps avoid both under-trust and over-trust. It is worth noting that the more controllable an AI system is, the stronger its tool-like attributes become, aligning more closely with the regulatory scope of traditional product liability frameworks. The greater the degree of independent decision-making and trust bestowed upon the technology, the greater the human dependence on it. Human cognitive frameworks may consequently become dominated by technology, with human–machine interaction increasingly resembling a *principal-agent* model. Conversely, when AI systems for clinical decision-making lack controllability due to their deep learning characteristics, human trust diminishes, and reflective verification of their decisions increases. In fact, this further reveals the limitations of *subject–object* dichotomy in understanding human–technology relations. Anthropocentrism and technocentrism often coexist. The more emphasis is placed on human control over technology, the greater the dependence on technology becomes, paradoxically subverting the subject–object relationship. The power dynamics between humans and technology are not characterised by mutual exclusion, but rather by a tense equilibrium that transcends unidirectional domination.

Thus, the central task is determining the consensus reached between technology providers and users regarding the trustworthiness of algorithms. The practice of the dialogic responsibility and related contractual agreements serve as crucial benchmarks. For example, consider a scenario where prior dialogues and agreements established that an AI system operates on rule-based algorithms and delivers outcomes consistent with current medical standards, yet the final diagnosis proves erroneous. In such a case, after assuming primary responsibility towards the patient, the medical institution may seek recourse from the technology provider. A potential concern arises: could technology providers circumvent liability by incorporating system alerts requiring medical personnel to independently review decisions? As previously established, the dialogic responsibility is

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 16 Nov 2026 at 18:59:47, subject to the Cambridge Core terms of use , available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/als.2026.10070

not a vague, blanket duty to warn. The technology provider must clearly specify the achievable accuracy level, identify which aspects require clinicians' verification, and clarify which aspects can be fully trusted. Otherwise, technology would not only fail to enhance clinical efficiency but become an additional burden. If the technology provider cannot make any commitment regarding freeing up clinicians' attention, clinicians would have no incentive to adopt such systems. Therefore, to achieve widespread adoption, technology providers naturally have an incentive to continuously refine usage guidelines and engage in communication with healthcare providers, thereby effectively addressing the preceding concerns.

Similarly, when employing deep learning-based systems, the generation of incomprehensible decisions does not equate to erroneous decisions violating medical standards. Rather, it represents an inherent probabilistic nature of the deep learning process, often accompanied by the system's ability to provide insightful guidance. This should not be regarded as a system flaw or a fault of the designers. In these circumstances, the scope of medical professionals' trust in AI systems should be bounded by the understandability of system decisions. Otherwise, reliance on system decisions would lose its fundamental rational basis and could no longer be termed *trust*. Therefore, if medical personnel, after reflecting on and verifying the diagnosis and treatment outcomes by integrating existing medical standards, their own clinical experience, the rationale and insights from the system's decisions, and the patient's actual condition, still deem the system's recommendations incomprehensible and choose not to adopt its conclusions, this does not constitute negligence on their part—even if future medical advancements validate the system's conclusions. Conversely, if medical personnel knowingly adopt the system's incomprehensible diagnostic or treatment decisions and cause harm, this constitutes negligence. In such a case, the medical institution and the healthcare provider shall bear liability for damages. It must be emphasised that, given that the vast databases of AI systems help medical personnel overcome temporal and spatial constraints to access medical information, their judgement on whether the system's decisions have a rational basis should be grounded in verification and reflection on the information provided by the system, rather than in their existing medical knowledge and experience alone.

Notably, regarding legal attributes of the dialogic responsibility in both the physician–patient relationship and the AI provider-user relationship, it should be clarified that a breach of dialogic responsibility does not automatically create a new, independent tort. Future statutes or binding standards could turn specified dialogue practices into direct duties of care. Until then, "relational responsibility" (and dialogic responsibility as its manifestation) is best understood as a principle for interpreting and coordinating existing tort, contract, professional, and regulatory duties. In the patient-facing relationship, it may serve as evidence that the institution and clinicians failed to meet the diagnostic and treatment duties under Article 1219 and Article 1221. In the provider–user relationship, failure to fulfil the dialogic responsibility may constitute breach of documented agreements and the good-faith duties of notification and assistance under Article 509, and such a breach may serve as a basis for internal recourse. In short, the introduction of relational responsibility aims to advance the relational interpretation of the law, thereby maximising its vitality. Relational responsibility thus does not dispense with positive law. Rather, it connects otherwise fragmented duties to the interaction that gives them practical meaning. It is also necessary to address a discussion in existing research: whether the development of clinical decision-making AI systems should incorporate medical personnel errors as a risk factor in system design. Existing research suggests that system developers should be able to anticipate the possibility of general negligence by medical personnel, but cannot be required to factor in gross negligence or even intentional misconduct (such as inaction in response to counter-intuitive machine judgements), as gross negligence constitutes an alternative cause that breaks the causal chain. This

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 16 Nov 2026 at 18:59:47, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/als.2026.10070

allocation of responsibility between humans and machines has two drawbacks: first, from a technical design perspective, the more a harmful outcome reflects gross negligence or even intentional misconduct by medical personnel, the more it indicates that the decision contradicts established medical knowledge and standards. Accordingly, this also implies that the decision-making process likely falls within an area where the AI system should have been capable of learning and providing accurate diagnostic conclusions. Conversely, matters that healthcare providers are prone to overlook due to ordinary negligence during re-evaluation may actually represent challenging aspects in algorithm design. This suggests that under the existing liability allocation standard in research, technical providers could be exempted from responsibility on the grounds that "the healthcare provider's gross negligence constitutes a substitute cause" precisely where the system likely harbours design flaws. This approach hinders the full realisation of the system's inherent decision-making capabilities, discourages technological refinement, and impedes improvements in accuracy. Second, discussions on whether "medical actions constitute a superseding cause" still implicitly follow the logic of linear causation and atomised attribution. This approach aligns with traditional product liability frameworks, limiting the provider's responsibility to foreseeable risks. However, traditional product liability could adopt this standard because a product's specific functions and production pathways are predetermined, providing clear guidance on subsequent usage and associated risks. If users act in ways exceeding expectations, this naturally constitutes an interruption of linear causation, thereby relieving producers of liability. However, in the context of AI applications for clinical decision-making, the relationship between system judgements and healthcare professionals' reviews is one of mutual reinforcement. Given this characteristic, the decision-making capacity of all parties—their control over outcomes—should be fully leveraged and facilitated through comprehensive interaction, aiming to achieve optimal clinical outcomes through collaborative efforts.

By contrast, the proposed framework offers two key advantages: first, it distinguishes between external responsibility aimed at fostering trust between physicians and patients and internal responsibility focused on enhancing algorithmic trust. The complexity of internal arrangements does not affect patient compensation, thereby safeguarding patient rights. For example, in scenarios where machine judgements are clearly counter-intuitive yet go unnoticed by physicians, crises of both physician–patient trust and algorithmic trust coexist. Patients have reasonable grounds to believe medical personnel failed in their duty, while physicians may reasonably question whether the AI system delivers its promised accuracy. This proposal effectively prevents mutual interference between these crises. Medical providers cannot evade liability by citing system provider errors but must assume primary responsibility before seeking recourse from the system provider based on their relative roles. This approach incentivises responsible action from both medical providers and technology suppliers. Second, the degree of trustworthiness can be determined through dialogue between technology providers and users, naturally incorporating market flexibility into the institutional design. Agreements reached through the active engagement of all parties can be embedded in the responsibility determination process, helping to avoid rigid application of the liability system. Thus, technological development and healthcare professionals' reflective verification become mutually reinforcing, providing feedback and refinement to each other.

5. Conclusion

Establishing responsibility systems is a socially constructed process requiring us to define the value foundations of "matters that should be borne" and set the direction for how people engage in cooperative interactions. The core reason traditional liability systems are

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 16 Nov 2026 at 18:59:47, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/als.2026.10070

built upon instrumental views of technology, linear causality, and atomised attribution frameworks lies in the long-standing dependence of social science research on the logic of natural science, attempting to use the latter's clarity to support the certainty and stability of institutions. However, this *descriptive attribution* obscures the *normative attribution* nature of institutions—namely, the underlying value presuppositions embedded within responsibility systems. Taking the debate over AI systems in medical decision-making as a case study, we can observe the limitations of a mechanistic worldview when confronting this issue. Building upon a reflection of existing institutional logic, this article proposes a theoretical framework of relational responsibility to guide the construction of concrete systems. Here, responsibility systems are not founded on the certainty of factual causation, nor do they primarily aim at punishing fault or compensating for damages. They also do not operate through atomised individual attribution. Instead, they fully acknowledge the complex interactions among all parties and the corresponding emergent outcomes rich in diverse possibilities. *Responsibility* signifies the commitment undertaken to build and develop wholesome relationships, emphasising a future-oriented grasp of the practical process, with *dialogic responsibility* as its core element. Specifically regarding the application of AI systems in clinical decision-making, the purpose of the responsibility system is to build trust in algorithms and between physicians and patients, thereby coordinating technological innovation, medical autonomy, and patient rights. Here, the trust relationship functions as a rational leap of faith, blending elements of *knowing* and *not knowing*. This characteristic aligns with the practical demands of operating within the realities of algorithmic black boxes and medical risks.

Based on this, the four points of contention in existing research can also be effectively addressed. First, AI systems for clinical decision-making do not possess inherent attributes but form a joint decision-making association with human decision-makers. Their decision-making patterns form a spectrum ranging from a near-agency relationship to a deep interactive relationship. Therefore, such AI systems cannot be confined within the binary framework of *product-service*. However, this does not mean that technology providers can evade their responsibilities. This article adopts a two-step strategy: first, it treats the human–machine association as the fundamental decision-making unit, encompassing all relevant actors and substantially reducing patients' evidentiary burden. Second, it designates medical institutions as the primary liability-bearing entity, while reserving to them the right to seek internal recourse against other decision-making participants. This balances liability while preventing mutual buck-passing and unduly harsh liability burdens. Second, this article acknowledges the *auxiliary role* of AI systems in clinical decision-making and advocates structuring responsibility arrangements based on this functional positioning. However, the interpretation of auxiliary role transcends a rigid, subject–object dichotomy in technology, avoiding the conflation of human–machine relationships with domination and preventing a static understanding of healthcare professionals' ultimate decision-making authority. In this article, the conception of auxiliary role signifies acknowledging AI's autonomous decision-making capabilities while adhering to the fundamental principle of *human-in-the-loop*. This ensures healthcare professionals undertake independent, reflective review of decisions, aiming for rule-based algorithms to precisely match rule-based clinical models and for deep learning algorithms to support heuristic clinical practice. Third, regarding the degree of duty of care for responsible parties, this article, grounded in the characteristics of the decision-making association, does not adhere to objective, static standards or atomised attribution schemes. Instead, it transforms this into an algorithmic trust issue. Specifically, it clarifies the level of trust healthcare professionals can place in specific AI systems. It advocates for dialogic responsibility between technology providers and users, determining whether the technology's design and use align with their reasonable expectations based on their dialogue and resulting agreements. This approach assesses each party's responsibility and

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 16 Nov 2026 at 18:59:47, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/als.2026.10070

whether they should be held accountable. Fourth, as previously stated, this article contends that the value of the responsibility system extends beyond mere compensation for damages, focusing instead on fostering beneficial relationships. Therefore, rather than having the human–machine decision-making association directly fulfil compensation responsibilities through mechanisms like independent funds, this article recognises the association as an abstract liable entity while still designating medical institutions as the primary centralised bearers of responsibility for claims brought by patients. This approach fosters trust between patients and healthcare providers while positioning medical institutions and practitioners as pivotal intermediaries. By leveraging physician–patient trust to empower algorithmic trust, it paves the way for societal acceptance and implementation of AI-driven clinical decision-making systems in the future.

It must be noted that this article does not rule out the possibility of AI systems directly engaging patients and earning their trust. However, realising this vision will require time. Achieving it requires building trust incrementally within the constraints of current societal realities, and can only be achieved once algorithmic trust becomes widespread. Accordingly, establishing independent funding mechanisms for the human–machine decision-making association will become increasingly important. This article does not reject such approaches but emphasises: premature reliance on them before society reaches that stage risks directly excluding key stakeholders—technology providers, users, and others—from active participation and responsibility-bearing. Under current societal conditions, institutional frameworks assigning responsibility to non-moral entities like AI would erode the sense of responsibility among moral agents, contradict present-day societal needs, and ultimately hinder technological advancement. Furthermore, to reconcile the tensions between technological innovation, medical autonomy, and patient rights, top-down technology licensing and regulation, along with the establishment of corresponding national and industry standards, are essential. However, the formation of top-down standards relies on the accumulation of bottom-up virtuous practices. This article focuses on constructing a responsibility system that regulates the interactive ecosystem among technology providers, medical practitioners, and patients. It offers a reflective, open, and inclusive institutional space for advancing practice, with the potential to accumulate experience for more mature systems in the future.

Acknowledgements. This article is an interim outcome of the Youth Fund Project of Humanities and Social Sciences Research of the Ministry of Education, entitled "Institutional Reflection on and Reconstruction of Family Elderly Care Agreements from the Perspective of Relational Jurisprudence" (25YJC820003). This research was also supported by the "Startup Program for Newly Appointed Young Faculty" of Shanghai Jiao Tong University (25X010505980).

References

AIGC Open Community (2025). *Bi 10 nian zhuanye yisheng zhun 4 bei! Weiruan fabu tupoxing yiliao AI xitong [Microsoft releases breakthrough medical AI system four times more accurate than a decade-experienced physician]*. Available at: https://mp.weixin.qq.com/s/iEtrYVmpxtRP2JLyKMGlpQ (Accessed: 25 February 2026).

Bao, A. and Zeng, Y. (2024). 'Understanding the dilemma of explainable artificial intelligence: A proposal for a ritual dialog framework', *Humanities and Social Sciences Communications*, 11, article no. 321.

Beckers, A. and Teubner, G. (2021). *Three liability regimes for artificial intelligence: Algorithmic actants, hybrids, crowds*. Oxford, New York: Hart Publishing.

Beckers, A. and Teubner, G. (2023). 'Human-algorithm hybrids as (quasi-)organizations? On the accountability of digital collective actors', *Journal of Law and Society*, 50(1), pp. 100–119.

Cai, D. (2023). 'Zhihui yiliao minshi zeren de neirong gexin yu peichang youhua [Content innovation and compensation optimization in civil liability for smart healthcare]', *Jianghan Tribune*, 12, pp. 121–126.

Center for Medical Device Standardization Administration NMPA (no date). *Medical device classification.* Available at: https://www.nifdc.org.cn/nifdc/xxgk/sjcx/ylqxfl/index.html (Accessed: 25 February 2026).

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 16 Nov 2026 at 18:59:47, subject to the Cambridge Core terms of use , available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/als.2026.10070

Chen, C. (2025). 'Rengong zhineng qinquan zhong de zeren zhuti ji guize yuanze [Liability subjects and attribution principles in artificial intelligence torts]', *Peking University Law Journal*, 37(2), pp. 363–380.
Cohen, I. G. (2020). 'Informed consent and medical artificial intelligence: What to tell the patient?', *Georgetown Law Journal*, 108(6), pp. 1425–1469.
Dai, W. and Wang, X. (2025). 'Suanfa touming jizhi de juxianxing jiqi kefu [The limitations of algorithmic transparency mechanisms and their overcoming]', *ECUPL Journal*, 28(1), pp. 33–44.
Dou, H. (2024). 'Zailun yiliao qinquan zhong yinguo guanxi de panduan [Revisiting the judgment of causation in medical torts]', *Nanjing University Law Journal*, 6, pp. 54–74.
Dou, H. (2025). 'Rengong zhineng chanpin qinquan guize heyi tuoli chanpin zeren [Why should liability for artificial intelligence product torts be separated from product liability]', *Peking University Law Journal*, 37(2), pp. 325–343.
Feinman, J. M. (2000). 'Relational contract theory in context', *Northwestern University Law Review*, 94(3), pp. 737–748.
Gergen, K. J. (2011). 'From moral autonomy to relational responsibility', *Zygon: Journal of Religion and Science*, 46(1), pp. 204–223.
Grasso, A. G. (2024). 'Diagnosi algoritmica errata e responsabilità medica', translated by Zhang, X., *Journal of Suzhou University (Law Edition)*, 11(4), pp. 154–166.
Heine, K. and Quintavalla, A. (2024). 'Bridging the accountability gap of artificial intelligence — What can be learned from Roman Law?', *Legal Studies*, 44(1), pp. 65–80.
Hou, M. (2025). 'Zhenliaoshi rengong zhineng de yiliao chanpin zeren rending [Determining medical product liability for diagnostic AI]', *Oriental Law*, 3, pp. 64–75.
Huang, Y. and Tan, Z. (2024). 'Shengchengshi rengong zhineng duiyu yihuan guanxi de lunli chonggou [Ethical reconstruction of doctor-patient relationships through generative artificial intelligence]', *Medicine and Philosophy*, 45(3), pp. 49–53.
Ji, H. (2020). 'Tayi guanxi yu jiqi zeren: Guanxilun de daode diwei guiyin ji yingyong [Alterity relations and machine's responsibility: A relational moral status ascription and its application]', *Studies in Dialectics of Nature*, 36(7), pp. 59–64.
Jiang, F. and Lin, J. (2024). 'Xushi yixue shijiao xia yihuan chongtu de chansheng yu zhili [The emergence and governance of doctor-patient conflicts from the perspective of narrative medicine]', *Chinese Medical Ethics*, 37(11), pp. 1325–1330.
Jones, C., Thornton, J. and Wyatt, J. C. (2023). 'Artificial intelligence and clinical decision support: Clinicians' perspectives on trust, trustworthiness, and liability', *Medical Law Review*, 31(4), pp. 501–520.
Li, R. (2020). 'Lun yiliao rengong zhineng de falv guizhi: Cong jinqi fangan dao yuanqi shexiang [On the legal regulation of medical artificial intelligence: From short-term plan to long-term plan]', *Administrative Law Review*, 4, pp. 46–57.
Li, R. and Shi, B. (2019). 'Rengong zhineng shiye xia yiliao sunhai zeren guize de shiyong he shanbian [The application and transformation of medical liability rules in the context of artificial intelligence]', *Journal of Shenzhen University (Humanities and Social Sciences)*, 36(6), pp. 91–99.
Li, X. (2016). 'Cong shizailun zouxiang guanxilun: Jiqiren lunli yanjiu de fangfalun zhuanhuan [From realism to relationism: A methodological transformation in roboethics]', *Studies in Dialectics of Nature*, 32(2), pp. 40–44.
Liang, L. (2023). 'Jiaohu yu gongxiang: Rengong zhineng yiliao juece de lunli xiangdu [Interaction and sharing: Ethical dimensions of AI medical decision-making]', *Medicine and Philosophy*, 44(10), pp. 19–24.
McNamee, S. and Gergen, K. J. (1999). *Relational responsibility: Resources for sustainable dialogue*. Thousan Oaks, London, New Delhi: Sage Publications.
Naiseh, M., Al-Thani, D., Jiang, N. and Ali, R. (2023). 'How the different explanation classes impact trust calibration: The case of clinical decision support systems', *International Journal of Human-Computer Studies*, 169, article no. 102941.
National Health Commission of the People's Republic of China (2025). Letter responding to proposal No. 04895 (healthcare category no. 408) from the third session of the 14th National Committee of the Chinese People's Political Consultative Conference. Available at: https://www.nhc.gov.cn/wjw/tia/202508/c989295636df4ce0b127141fb5c2dddc.shtml (Accessed: 25 February 2026).
National People's Congress of the People's Republic of China (2020). Civil Code of the People's Republic of China. Available at: https://www.court.gov.cn/zixun/xiangqing/233181.html (Accessed: 21 July 2026).
Painter-Morland, M. (2006). 'Redefining accountability as relational responsiveness', *Journal of Business Ethics*, 66(1), pp. 89–98.
Shumway, D. and Hartman, H. (2024). 'Medical malpractice liability in large language model artificial intelligence: Legal review and policy recommendations', *Journal of Osteopathic Medicine*, 124(7). pp. 287–290.
Simmel, G. (2002). *Soziologie. Untersuchungen über die formen der vergesellschaftung [Sociology: Inquiries into the construction of social forms]*. Translated by Lin, R. Beijing: Huaxia Publishing House.
Terry, N. (2019). 'Of regulating healthcare AI and robots', *Yale Journal of Law and Technology*, 21(3), pp. 133–190.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 16 Nov 2026 at 18:59:47, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/als.2026.10070

Unver, M. B. (2023). 'Governing fiduciary relationships or building up a governance model for trust in AI? Review of healthcare as a socio-technical system', *International Review of Law, Computers & Technology*, 37(2), pp. 198–226.

Wang, L. (2024). 'Rengong zhinengti de fuze kunjing jiqi guanxishi chulu xinjie [The accountability dilemma of artificial intelligence entities and a new relational approach to resolution]', *Journal of Northeastern University (Social Sciences)*, 26(5), pp. 15–22.

Wu, X., Huang, T., Deng, L., Qiao, Y., Razzak, I. and Xie, Y. (2025). 'A knowledge-driven adaptive collaboration of LLMs for enhancing medical decision-making', *2025 Conference on Empirical Methods in Natural Language Processing*. Suzhou: Association for Computational Linguistics, pp. 33495–33512.

Yang, J., Shu, L., Duan, H. and Li, H. (2025). 'RDguru: A conversational intelligent agent for rare diseases', *IEEE J Biomed Health Inform*, 29(9), pp. 6366–6378.

Zheng, Z. (2023). 'Zhenliao rengong zhineng de yiliao sunhai zeren [Medical liability for diagnostic AI]', *China Legal Science*, 1, pp. 203–221.

Cite this article: Chen X (2026). Establishing Responsibility Systems for AI Application in Clinical Decision-Making: A Relational Responsibility Perspective. *Asian Journal of Law and Society* **13**, 630–653. https://doi.org/10.1017/als.2026.10070

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 16 Nov 2026 at 18:59:47, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/als.2026.10070