Computers and Artificial Intelligence；Vol.3, No.2; 2026
E-ISSN: 3006 -3728 P-ISSN：3006-371X
DOI: https://doi.org/10.70267/cai.26v3n2.2834
Published by: Zeus Press

Responsibility Definition and Risk Management in the Clinical Application of Medical Artificial Intelligence: A Review Based on Four-Level Classification

Shihan Yin*

School of Management, Tianjin University of Technology, Tianjin, China

**Corresponding author: Shihan Yin.*

Abstract

The rapid penetration of medical artificial intelligence (MAI) into clinical diagnosis and treatment scenarios has reshaped the traditional medical service model. However, issues such as ambiguous responsibility definition and lagging risk management have severely constrained its safe and compliant development. This article uses the four-level AI classification system (tool-type, advisor-type, collaborative-type, autonomous-type) as the analytical framework to systematically review the responsibility definition logic and full-lifecycle risk management mechanisms for MAI clinical applications. The study finds that a multi-stakeholder responsibility system covering manufacturers, medical institutions, doctors, and regulatory authorities has been formed, along with a "prevention-control-remediation" risk management closed loop. However, deficiencies still exist in empirical validation, standard unification, adaptation to special scenarios, and coordination between technology and institutions. Future research should focus on directions such as dynamic responsibility quantification and matching, practical design of insurance pools, and development of lightweight management tools for primary care, providing theoretical support and practical references for the systematic governance of MAI clinical applications.

Keywords

medical artificial intelligence, responsibility definition, risk management, four-level classification system, dynamic responsibility matrix, full-lifecycle governance

1. Introduction

In recent years, medical artificial intelligence (MAI) has been deeply integrated into clinical processes such as disease screening, diagnosis, treatment, and health management. With its efficient data processing capabilities and precise decision-support functions, it has become an important tool for optimizing medical resource allocation and improving diagnostic and treatment efficiency. However, risks hidden behind technological innovation—such as algorithmic black boxes, ambiguous responsibility, and data privacy breaches—are becoming increasingly prominent. Responsibility attribution disputes in human-machine collaborative models and full-process risk management dilemmas have become the core bottlenecks restricting MAI from moving from technological implementation to compliant application [2].

The existing MAI clinical application governance system faces three core challenges. First, responsibility definition lacks a unified logic, with blurred boundaries of rights and obligations among multiple stakeholders, making it difficult to adapt to the clinical application scenarios of AI products with varying degrees of autonomy currently on the market. Second, risk management processes are fragmented, with a lack of systematic linkage among data governance, algorithm regulation, emergency response, and other segments. Third, existing norms are disconnected from clinical practice, showing insufficient adaptability to special scenarios such as primary care and rare disease treatment [4]. These issues not only affect medical quality and patient safety but also hinder the sustainable development of MAI technology.

This article introduces the four-level AI classification system proposed by Zhang et al. (2026) as the core analytical framework. Based on the dual dimensions of "clinical decision-making autonomy" and "impact on doctor-patient relationships," this system divides medical AI into four categories: tool-type, advisor-type, collaborative-type, and autonomous-type [3], providing a scientific basis for differentiated responsibility allocation and risk management. The article focuses on the two core themes of "responsibility definition" and "risk management processes," systematically integrates core domestic and international literature, policy documents, and expert consensuses [1][3], and aims to construct a logically clear and highly practical governance framework to provide reference for policymakers, medical institutions, technology developers, and the healthy development of the field.

2. Logical Framework for Responsibility Definition

2.1 Multi-Stakeholder Responsibility System

Responsibility definition for MAI clinical applications involves four core stakeholders. The rights and obligations of each stakeholder are both independent and collaboratively linked, jointly forming a complete responsibility chain:

Technology manufacturers, as the primary developers of medical artificial intelligence, bear source-level technical responsibility. Their core duties include ensuring the safety and reliability of model algorithms, providing users with complete technical documentation and risk warnings, and ensuring the compliance of product iterations. They bear primary responsibility for medical risks caused by algorithmic defects or data biases [1].

For medical institutions, they should fulfill management and supervisory responsibilities, including adapting the MAI deployment environment, standardizing usage procedures, and conducting quality control and risk monitoring. They bear management responsibility for risks arising from improper system integration or non-standardized operational procedures [3].

Clinical doctors, as the ultimate responsible parties for diagnostic and treatment decisions, bear responsibility for clinical review, risk identification, and patient communication. They must professionally judge MAI output results and make personalized adjustments based on the patient's own conditions, while fulfilling necessary information disclosure obligations [7].

Regulatory authorities shoulder the responsibility of policy formulation and supervision. They are responsible for establishing a classified and graded regulatory system, improving admission review standards, and conducting dynamic oversight. They bear regulatory responsibility for systemic risks caused by lagging regulatory standards or ineffective supervision.

2.1.1 Comparative Analysis

The responsibility boundaries of different stakeholders exhibit a layered characteristic of "technology end–application end–regulatory end," yet scholarly understanding of the weight of each stakeholder's responsibility shows differences. Some studies (e.g., [1]) tend to assign primary responsibility to technology manufacturers, arguing that algorithmic defects and data biases are the root causes of risk. Other studies (e.g., [3]) emphasize the process management responsibility of medical institutions, believing that improper deployment and non-standardized procedures are the key factors amplifying risk. Still other studies (e.g., [7]) focus on doctors' ultimate decision-making responsibility, asserting that the core position of clinical judgment is irreplaceable. In international research, the EU Artificial Intelligence Act places greater

emphasis on the risk-classified regulatory responsibility of regulatory authorities, while WHO guidelines balance manufacturers' technical responsibility with doctors' supervisory responsibility [3]. However, in multi-platform systems, technological collaboration and complementarity among different platforms constitute another major advantage. Platforms complement each other in areas such as drug design and immune response enhancement, making vaccine design more comprehensive and precise [5].

2.2 Dynamic Responsibility Matrix Based on Four-Level Classification

Combining the degree of AI autonomous decision-making with clinical risk scenarios, a differentiated dynamic responsibility matrix is constructed. The specific allocation ratios are shown in Table 1:

Table 1: Dynamic Responsibility Allocation Chart

AI Type	Core Responsible Stakeholder	Responsibility Allocation Ratio	Core Responsibility
Tool-type	Clinical doctor	Doctor 80%+, Manufacturer 20%-	Doctor responsible for information review and clinical application; manufacturer provides technical support
Advisor-type	Doctor + Manufacturer	Doctor 50%, Manufacturer 50%	Doctor responsible for professional review and decision-making; manufacturer bears responsibility for algorithmic defects
Collaborative-type	Medical institution + Manufacturer	Institution + Manufacturer 60%, Doctor 40%	Institution ensures procedural compliance; manufacturer ensures algorithm transparency; doctor responsible for process supervision
Autonomous-type	Manufacturer	Manufacturer 60%-80%, Doctor 20%-40%	Manufacturer ensures system operation within authorized scope; doctor fulfills supervision and intervention responsibilities

This matrix is clearly defined through formal agreements and regulatory documents. It is dynamically adjusted in conjunction with clinical scenario risk levels and accident investigation results, avoiding a "one-size-fits-all" responsibility allocation model and achieving precise matching of responsibility with AI autonomy and risk levels [1][3].

2.3 Core Difficulties in Responsibility Determination

Responsibility determination for MAI clinical applications faces multiple challenges. The core difficulties concentrate on four aspects: the algorithmic black-box dilemma, blurred human-machine boundaries, overlapping multi-factor causation, and disputes over subject qualification. The algorithmic black box arises because the decision-making process of deep learning models lacks transparency, making it difficult to trace the root cause of errors and resulting in a lack of clear basis for technical responsibility definition [4]. The decision-making boundaries between MAI and doctors are intertwined; when diagnostic and treatment outcomes deviate, it is difficult to distinguish whether the cause is algorithmic defects, operational errors, or clinical judgment deviations, leading to blurred human-machine boundaries [7]. Patient harm is often caused jointly by multiple factors such as data bias, algorithm drift, and doctor operations, resulting in complex causal chains and difficulty in quantifying responsibility proportions [3]. Moreover, the academic community currently holds divergent views on whether AI possesses legal personality. The mainstream view treats it as an advanced medical tool, yet the dual judgment risks of AI and clinical doctors make responsibility tracing even more complex [1].

3. Full-Process Governance of Risk Management

3.1 Risk Prevention: Source Governance

Risk prevention, as the primary link in risk management for medical artificial intelligence (MAI) clinical applications, focuses on three core segments—data, algorithms, and pre-deployment validation—to build a comprehensive source-level control system. At present, the Hangzhou Institute of Medical Research of the Chinese Academy of Sciences relies on the professional knowledge of its interdisciplinary teams, combining clinical data and cutting-edge algorithms, to continuously evaluate vaccine performance in the experimental stage; it also conducts in-depth risk assessments from pre-clinical data through to the clinical trial stage,

ensuring that every link in vaccine design, production process, and clinical application meets strict quality standards and regulatory requirements [5].

In terms of data governance, the core principles of "reliability, privacy protection, and bias control" are applied. The training data for medical artificial intelligence are required to fully cover local high-incidence diseases and special populations. Privacy computing technologies such as federated learning are adopted to implement the "data stays in place, model moves" application model [9]. At the same time, a comprehensive data bias assessment and correction mechanism is established to prevent algorithmic discrimination at the source [2].

In terms of algorithm governance, a dual regulatory system of ethical certification and third-party auditing is implemented. High-risk medical artificial intelligence products must pass both medical device approval standards and clinical adaptability assessments [1]. Independent third-party professional institutions conduct regular audits of algorithm fairness, safety, and transparency, with audit results required to be disclosed to the public and subject to industry and public supervision [3].

In terms of pre-deployment validation, a stepped implementation process of "low-risk pilot → multi-center validation → full-scenario promotion" is promoted. Prospective clinical trials are conducted, with trial coverage required to include different regions, different levels of medical institutions, and special clinical populations [4]. A dual validation system of "technical indicators–clinical value" is also established to ensure both technical feasibility and clinical practicality of medical artificial intelligence products [1].

3.2 Risk Control: Dynamic Intervention

Risk control focuses on the entire process of medical artificial intelligence clinical application and achieves closed-loop risk management through real-time monitoring, threshold constraints, rapid response, and dynamic calibration measures [3].

In usage monitoring, a confidence labeling and conflict early-warning mechanism is established. When medical artificial intelligence outputs diagnostic and treatment results, it must simultaneously provide quantitative confidence scores and decision reasoning basis [1]. The system also compares algorithm conclusions in real time with the patient's medical history and test results; when significant inconsistencies appear, an immediate warning is triggered to remind clinicians to intervene and judge [4].

In threshold control, differentiated operating thresholds are set based on the type of medical artificial intelligence and the risk level of the clinical application scenario. For high-risk clinical scenarios, the algorithm's hallucination rate must be controlled at $\leq 3\%$; for low-risk scenarios, it may be appropriately relaxed to $\leq 5\%$ [3]. For autonomous-type medical artificial intelligence, clear authorization thresholds and application boundaries for independent decision-making must be defined to prevent out-of-scope use [1].

In the emergency fuse mechanism, when medical artificial intelligence experiences consecutive major misdiagnoses, serious safety hazards, or system operation anomalies, its clinical application must be immediately suspended [2]. Complete system operation logs must be preserved, and a third-party professional institution must conduct a special algorithm audit. Clinical use may only be resumed after problem rectification is completed and re-inspection is passed [3].

In continuous learning monitoring, core performance indicators after model iteration and updates are tracked in real time. Clear performance degradation thresholds are set; when the algorithm misdiagnosis rate rises or core performance declines by more than 10%, an immediate manual review and version rollback procedure is triggered [4] to ensure that model iterations do not introduce new systemic biases.

3.3 Risk Remediation: Post-Incident Safeguards

Risk remediation builds a complete chain of "traceability–compensation–accountability" to provide post-incident protection for patient rights and interests:

In post-incident traceability, each MAI system is assigned a unique full-lifecycle traceability code. Blockchain technology is used to record key information such as model versions, operation logs, and iteration history, achieving full-process visualization and traceability of decisions and providing a basis for responsibility determination [1].

In economic compensation, the government guides manufacturers and medical institutions to jointly establish a medical AI accident insurance pool [2]. Compensation is provided for damages caused by systemic risks during compliant algorithm iterations. A fast claims channel is established, and compensation ratios for different AI types are clearly defined [3].

In accountability, relevant domestic legislation and regulations on artificial intelligence do not treat it as a separate legal personality but define it as a "medical device." The Medical Device Classification Catalogue was issued in 2017, and the revised detailed rules specifically added the minor category of "artificial intelligence-assisted diagnosis and treatment" [6]. A graded accountability mechanism is established; based on the proportion of fault of the responsible party and the severity of the damage consequences, graduated measures such as economic penalties, qualification suspension, and industry bans are implemented. Doctors' AI usage qualifications and responsibility fulfillment are incorporated into performance assessments [2].

3.4 Existing Problems in Risk Management

The current risk management system still has four prominent problems: First, industry standards are not unified; significant differences exist in medical artificial intelligence data annotation norms, operation monitoring indicators, and threshold standards across different R&D institutions and medical institutions, lacking industry-recognized quantitative norms and severely affecting governance consistency [4]. Second, empirical validation is insufficient; most control mechanisms remain at the level of expert consensus and theoretical frameworks, lacking large-scale clinical practice validation, and actual operational effectiveness remains unknown [1]. Third, primary medical institutions lack lightweight monitoring tools adapted to limited computing power; doctors' risk identification and control capabilities are inadequate, making it difficult to meet primary care scenario needs [2]. Fourth, the synergy mechanisms between technologies such as algorithm explainability and blockchain traceability and institutions such as responsibility determination and data governance are imperfect, failing to form joint governance forces [3].

4. Conclusion and Outlook

4.1 Main Research Findings

This article uses the four-level AI classification system as the framework to systematically review the responsibility definition and risk management logic for MAI clinical applications. The core findings are as follows:

In responsibility definition, medical artificial intelligence clinical applications have formed a multi-stakeholder responsibility system covering technology manufacturers, medical institutions, clinical doctors, and regulatory authorities. The dynamic responsibility matrix constructed on the basis of AI autonomous decision-making degree achieves precise matching of responsibility allocation with AI type and clinical risk level, providing a scientific tool for differentiated governance of MAI clinical applications.

In risk management, a full-lifecycle risk management process of "prevention–control–remediation" has been constructed. Core mechanisms such as data governance, ethical certification, threshold control, and the medical AI accident insurance pool are interconnected and synergistic, forming a systematic risk management system that provides institutional safeguards for the safe and compliant operation of MAI clinical applications.

In governance logic, the four-level classification system runs through the entire process of responsibility definition and risk management for medical artificial intelligence clinical applications, achieving deep synergy of "AI type–responsibility allocation–risk management," effectively solving the fragmentation problems in traditional governance models and providing an effective path for optimizing the MAI clinical application governance system.

4.2 Research Limitations

Although this study systematically reviews responsibility definition and risk management for MAI clinical applications, it still has four obvious research shortcomings: First, empirical research is relatively scarce; most governance mechanisms remain at the level of theoretical discussion and expert consensus, and

core mechanisms such as the dynamic responsibility matrix and medical AI accident insurance pool lack support and validation from large-scale clinical data regarding their actual operational effectiveness. Second, industry standards and norms are missing; key elements such as responsibility proportion quantification, risk threshold control standards, and third-party algorithm audit norms have not yet formed unified industry standards, resulting in insufficient practicality of governance mechanisms. Third, research on adaptation to special scenarios is insufficient; governance solutions targeting special clinical scenarios such as primary care, rare disease diagnosis and treatment, and integrated traditional Chinese and Western medicine are relatively few, making it difficult to meet diversified clinical application needs. Fourth, research on technology–institution synergy is not deep enough; the linkage between algorithm technology R&D and legal norms or regulatory policies is not sufficiently close, failing to fully realize benign interaction between technology empowerment and institutional constraints.

4.3 Future Research Directions

Combining the conclusions and limitations of this study, future research on governance of medical artificial intelligence clinical applications can focus on the following key directions for deeper exploration:

First, advance quantitative matching research on dynamic responsibility. Based on risk levels of different clinical scenarios and AI decision-making participation, establish quantitative models of responsibility proportions and control measures to further improve the precision and scientific nature of MAI clinical application governance.

Second, conduct practical design and pilot application of medical AI accident insurance pools. Clarify key issues such as funding contribution ratios, compensation standards, and claims processes for the insurance pool, and continuously optimize the mechanism design through pilot applications in different regions and different levels of medical institutions.

Third, develop lightweight risk management tools for primary care scenarios. Fully adapt to the real conditions of insufficient computing power and shortage of professional personnel in primary medical institutions, while strengthening training for primary clinical doctors to comprehensively enhance risk management capabilities for medical artificial intelligence in primary care.

Fourth, strengthen localized adaptation research on international governance experience. Draw on advanced international experiences such as the EU Artificial Intelligence Act’s risk-classified regulatory model and the World Health Organization’s human autonomy principles, and optimize the governance framework for MAI clinical applications in combination with the development characteristics of China’s medical system and actual clinical needs.

Fifth, reinforce synergistic innovation research between technology and institutions. Promote deep linkage between core technologies such as algorithm explainability, privacy computing, and blockchain traceability and institutional construction such as responsibility determination, data governance, and industry regulation, achieving synchronized resonance between technology R&D and institutional design to enhance governance effectiveness for MAI clinical applications.

References

[1] Gong, M. C., Ma, Y. H., Pan, H., Bai, H., Dai, H., Chen, W., ... & Ji, X. M. (2025). Expert consensus on ethical governance of clinical applications of generative medical artificial intelligence (2025 edition). *Acta Academiae Medicinae Sinicae*, 1-14.

[2] Gong, M. C., Li, Y. H., Ma, Y. H., Gong, K., Liu, C., Ouyang, Z. H., & Dai, H. (2026). Ethical governance of generative medical artificial intelligence: Three-dimensional collaborative path and Chinese practice. *Journal of Medical Informatics*, 47 (01), 2-8+23.

[3] Zhang, C. (2026). The legal dilemma of medical artificial intelligence in China: challenges to physicians' duty to inform and a typology-based response. *Frontiers in Public Health*, 13, 1747635-1747635. https://doi.org/10.3389/FPUBH.2025.1747635.

[4] Cheng, W. M., & Li, G. M. (2025). Medical artificial intelligence: From technical performance to clinical utility. *Guangdong Medical Journal*, 46 (11), 1601-1605. https://doi.org/10.13820/j.cnki.gdyx.20253316.

[5] He, C. L., & Qiu, R. (2025). AI empowering biomedicine to reshape the future of life sciences—Entering the medical artificial intelligence center of the Hangzhou Institute of Medicine, Chinese Academy of Sciences. *High Technology and Industrialization*, 31 (11), 16-18. https://doi.org/10.26927/j.cnki.hitech.2025.11.004.

[6] Huang, L. N., & Zhang, L. (2025). Exploration of legal responsibility subjects in tortious acts of artificial intelligence medical products. *Journal of Health Law*, 33 (06), 32-42. https://doi.org/10.19752/j.cnki.2097-5058.2025.06.004.

[7] Duffourc, M. & Gerke, S. (2023). Generative AI in Health Care and Liability Risks for Physicians and Safety Concerns for Patients.. JAMA, 330 (4), https://doi.org/10.1001/JAMA.2023.9630.

[8] Luo, S. n., & Wang, H. P. (2022). Advances in the application of artificial intelligence in the field of emergency nursing. *Nursing Research*, 36 (05), 884-887.

[9] Sheller, M. J., Edwards, B., Reina, G. A., Martin, J., Pati, S., Kotrotsou, A.... & Bakas, S. (2020). Federated learning in medicine: facilitating multi-institutional collaborations without sharing patient data.. *Scientific Reports*, 10 (1), 12598. https://doi.org/10.1038/s41598-020-69250-1.

Funding

This research received no external funding.

Conflicts of Interest

The authors declare no conflict of interest.

Acknowledgment

This paper is an output of the science project.

Copyrights