Embodied Artificial Intelligence in Healthcare: Foundations, Applications and Future Challenges

Jinyang Li*

Information Management and Information Systems, Anhui University, 230601 Hefei, China

Abstract. This study aims to systematically review the current applications, technological foundations, and deep-seated challenges of Embodied Artificial Intelligence (EmAI) in healthcare. It validates EmAI's immense potential to drive the transformation of medical intelligence from "screen-based intelligence" to "scenario-based intelligence," which is crucial for building next-generation smart healthcare systems. The methodology involves a systematic review focusing on summarizing EmAI's performance in clinical interventions and assisted living, core technologies e.g. multimodal perception, large-model reasoning, and world models, as well as the urgent need for federated learning and responsible research and innovation principles. Quantitative findings confirm EmAI has evolved into an intelligent collaborator rather than a mere auxiliary tool: AI-assisted surgery improved surgical precision by 40% while significantly reducing complications and operating time. Within assistive technologies, EmAI substantially empowered the independence of visually impaired individuals, embodying a "human-centered" design philosophy. These conclusions support the core argument that EmAI is a key driver of healthcare transformation. Future research should focus on advancing FL to address data generalization and enhance the real-time decision-making robustness of world models. Prioritizing the establishment of clear ethical, legal, and fairness frameworks is essential to ensure EmAI evolves into a safe, reliable, and universally accessible force for healthcare transformation.

1 Introduction

Since the inception of digital computing concepts, Artificial Intelligence (AI) has carried the grand ambition of achieving Artificial General Intelligence (AGI). Embodied Intelligence (EI) is widely regarded as one of the key pathways to realizing this goal. EI emphasizes the deep integration of "physical entities" (such as robots and wearable devices) with "agents" (AI algorithms and models) to achieve dynamic, real-time perception-action-cognition fusion with the physical world. This enables autonomous learning and adaptive evolution [1]. With the maturation of cutting-edge AI technologies like multimodal large language models (MLLMs) and "world models," embodied intelligence demonstrates immense potential for

* Corresponding author's email: l12314106@stu.ahu.edu.cn

transformative applications in critical domains such as smart manufacturing, home care, and intelligent healthcare.

Global healthcare systems face systemic challenges including population aging, increasing burdens of chronic diseases, and uneven distribution of medical resources. Traditional “screen-based AI,” while demonstrating high accuracy in disease prediction and medical imaging analysis [2], remains limited in its impact due to a lack of deep physical integration with real clinical or home environments and the ability to execute actions [1]. It cannot proactively operate surgical instruments, assist patient mobility, or perform care tasks in unstructured home settings. It also struggles to effectively address “hallucinations” and task generalization challenges in physical interactions. Existing assistive technologies and early robotic applications (e.g., remote surgical assistance) have made progress in specific tasks, yet their core limitations lie in the absence of true real-time decision autonomy, complex situational awareness, and natural human-machine interaction capabilities. These pre-programmed systems struggle to adaptively function in dynamic, unstructured environments [3]. Furthermore, home deployment faces complex challenges including maintenance, user privacy, and long-term security [4].

Therefore, this research aims to address the core challenge: how to translate AI's powerful cognitive capabilities into safe, reliable, and efficient physical action, achieving a systemic leap from “AI on screens” to “embodied systems in scenes.” This is a necessary prerequisite for advancing next-generation smart healthcare technologies, enhancing independence for individuals with disabilities, and improving rehabilitation efficiency [5,6]. It also fills the critical gap in existing technologies' execution within the physical world.

The application of embodied intelligence in healthcare is rapidly advancing, primarily manifested in three areas:

Clinical Intervention and Rehabilitation Therapy: AI-driven surgical systems have been rapidly adopted in fields such as orthopedics and general surgery. Through real-time image segmentation (e.g., 3D CT image segmentation models [7]) and path planning, they significantly enhance surgical precision and reduce procedure duration [3]. In rehabilitation, wearable robotic devices (e.g., exoskeletons) are integrating embodied perception, interaction, and cognitive models to achieve precise recognition of human movement intent and physical function restoration.

Mobility Assistance and Independent Living: Research has evolved from basic mobility aids to integrated advanced intelligence, addressing challenges in mobility, information access, and social interaction for individuals with disabilities. Embodied AI wheelchair frameworks combine hands-free interfaces (e.g., head tracking, voice commands) with advanced autonomous navigation algorithms, enabling greater independence for individuals with severe disabilities [5]. For frailty care among the elderly, home-based assistive robots are being explored to provide continuous support, though questions remain regarding user privacy protection and achieving safe, efficient human-robot collaboration [4].

Ethical and Social Dimensions: As embodied AI permeates assisted living and care, systematic discussions on its implications for bioethics, social equity (i.e., “Disability 4.0” issues) [4,7], and human-robot interaction safety have become increasingly urgent [8, 9].

However, existing research reviews often focus on specific technologies or single application scenarios, lacking a unified analytical framework to comprehensively map the key technologies underpinning these diverse applications. This review aims to fill this gap by providing a comprehensive, cross-domain systematic analysis of embodied intelligence in healthcare applications. This paper not only summarizes major research findings but also constructs a unified “technology-application-challenge” analytical framework. This framework highlights how embodied intelligence overcomes limitations of traditional AI in medical physical interactions through the closed-loop integration of perception-action-cognition, while offering forward-looking guidance on critical challenges such as data generalization, real-time robustness, and bioethical safety.

The remainder of this paper is organized as follows: Section 2 outlines the core technological foundations of embodied intelligence in healthcare (multimodal large models, world models, embodied perception and decision-making mechanisms). Section 3 systematically reviews recent advancements in three major application scenarios of embodied intelligence in healthcare: assisted mobility and independent living, wearable rehabilitation robots, and AI-assisted autonomous surgery. Section 4 delves into key challenges embodied AI faces in medical settings (data generalization, real-time robustness, bioethics and safety). Section 5 concludes the paper and outlines potential research directions for embodied intelligence in building next-generation intelligent healthcare systems.

2 Method

This study conducts a systematic analysis of the application of Embodied Artificial Intelligence (EmAI) in the healthcare sector. The core methodology involves the generalization, synthesis, and interpretation of the materials, preparation procedures, system architectures, and evaluation strategies employed in the fifteen reviewed literatures. This section systematically refines these research methods to construct the core technical foundation of medical Embodied AI systems, including hardware platform design, and the realization paths of key algorithms such as multi-modal large models, world models, and embodied perception and decision-making mechanisms.

Through chronological and causal analysis, this paper analyzed the evolutionary process of Embodied AI at the levels of physical hardware platforms, core AI algorithms and models, and evaluation strategies. This paper categorized these research methods into two main themes: System Development and Algorithm Construction, and Evaluation and Analysis Strategies, ensuring a clear presentation of the technical basis in this field. Specific application scenarios (such as assistive mobility, wearable rehabilitation robotics, and AI-assisted autonomous surgery) are realized and evaluated based on these technical methods.

2.1 Hardware and Platform Development of Embodied AI Systems

The core of most Embodied AI research lies in designing and manufacturing a physical platform capable of perceiving and interacting with the physical world. The reviewed literature details the complex hardware, sensing materials, and integration preparation processes required to build Embodied AI for different application scenarios.

2.1.1 Materials and Preparation for Mobility and Assistive Platforms

In assistive mobility solutions, researchers focus on advanced Embodied AI retrofitting of existing wheelchairs and mobility platforms to achieve autonomous navigation and hands-free operation. For instance, the Embodied AI wheelchair frame designed for severely disabled individuals includes a custom electric wheelchair chassis, a robotic arm, a retractable canopy, and a comprehensive sensor suite for environmental data acquisition [5]. Sensor materials are crucial for Simultaneous Localization and Mapping (SLAM) and collision avoidance, including multi-line Lidar units, depth-sensing cameras (e.g., RealSense D435i), and Inertial Measurement Units (IMUs), designed to collect high-resolution distance and visual data. The preparation process involves the mechanical installation, electrical connection, and precise calibration of these sensors, particularly to ensure that Lidar and camera data can undergo multi-modal fusion to generate accurate 3D environmental maps.

Furthermore, the human-machine interface materials for this framework include a head-tracking system and voice control module for user command input.

2.1.2 Human-Robot Interaction Materials for Wearable Robotics

In the field of wearable robotics (e.g., powered exoskeletons and orthoses), researchers use specialized actuators and a range of physiological signal acquisition devices as core materials [6]. The material list typically involves high torque density motors, reducers, and sensing signal materials used to identify human movement intent and physiological states. These materials include surface electromyography (sEMG) sensors, force/torque sensors, and kinematic sensors for measuring joint angles and limb kinematics. Preparation work focuses on the mechanical and electrical integration of these components to enable robust physical interaction control (based on interaction forces) and human-in-the-loop control (relying on human physiological or biomechanical feedback signals). The research emphasizes that this integration method aims to enable the robot to understand and execute human movement intentions, forming a closed-loop perception-action feedback circuit.

2.2 Core AI Algorithm and Model Construction Methods

The functional realization of Embodied AI relies on complex computational materials and model construction methods. These methods primarily involve the selection of deep learning models, preparation of training data, and application of optimization strategies to endow the embodied system with multi-modal perception, world understanding, and action decision-making capabilities.

2.2.1 Technical Roadmap for Medical Image 3D Segmentation Models

For surgical and diagnostic Embodied AI applications, such as assistive surgical robots, researchers focused on analyzing and preparing three main categories of deep learning models [7]. These model types include: Convolutional Neural Network (CNN)-based models (e.g., 3D U-Net variants), Transformer-based models (e.g., nnFormer), and extensions of Segment Anything Model (SAM) for universal segmentation. Material preparation involves training models on large-scale medical CT image datasets to perform accurate three-dimensional (3D) segmentation of anatomical structures. The training process (preparation work) adopts optimization strategies to address key challenges in clinical segmentation tasks, including loss function optimization, data augmentation, and multi-modal fusion. For example, the development of the nnFormer model focuses on using the Transformer architecture to capture global contextual information, addressing 3D spatial consistency and small-sample adaptability issues, thereby providing precise anatomical guidance for the embodied system.

2.2.2 Training Process for Autonomous Navigation and Decision Mechanisms

In mobile assistive platforms, the core of the software preparation is a novel self-learning AI loop [5]. This method aims to achieve real-time autonomous navigation and avoid stationary and dynamic objects. Training materials are derived from environmental data and 3D maps collected by sensors. The model utilizes mechanisms such as reinforcement learning or imitation learning for iterative optimization in virtual or real environments. The key to this construction process is its emphasis on the robustness of real-time decision-making, ensuring that the Embodied AI can safely and effectively execute tasks in dynamic, unstructured real-world environments, such such as home settings or hospital corridors, reflecting an implicit understanding and application of a "world model."

2.3 Evaluation, Analysis, and Statistical Methods for Embodied AI

A comprehensive assessment of the impact of Embodied AI in healthcare requires the use of multiple evaluation methods and statistical tests, including quantitative, qualitative, and systematic review approaches.

2.3.1 Systematic Review and Quantitative Analysis of Clinical Efficacy

For the clinical application of AI-assisted surgery, researchers employed systematic review and meta-analysis statistical testing methods [3]. The methodological process involved extracting quantitative data from 25 recent peer-reviewed studies (2024-2025) to evaluate the outcomes of AI-assisted robotics relative to traditional manual surgical techniques in terms of clinical efficacy, surgical precision, complication rates, and economic benefits. The focus of the statistical tests was on comparing the average values of surgical time, percentage reduction in intraoperative complications, percentage improvement in surgical precision, and reduction in patient recovery time, thereby providing evidence-based medical evidence for the application of AI in the surgical field.

2.3.2 Measurement Methods for User Functionality and Satisfaction

In the field of assistive technology, such as Embodied AI prototypes for the visually impaired, researchers adopted long-term observational studies to measure functional improvement [10]. Measurement methods included observation and data collection of participants before and after use over a period of 60 days. Statistical tests and measurement tools primarily rely on the Visual Functioning Subscale (VFS) to quantitatively assess improvements in navigation, safety, and social participation provided by the prototype system. This method emphasizes the practical impact of the embodied system on the user's independent living ability and quality of life, validating the effectiveness of the technological intervention by comparing VFS scores before and after use.

2.3.3 Qualitative Analysis Framework for Policy, Ethics, and Challenges

Some studies adopted a qualitative analysis framework to explore the non-technical challenges faced by Embodied AI. Analysis methods typically involve literature review and thematic analysis. For instance, some studies analyzed challenges faced when integrating assistive technology into homes for the aging population, such as affordability, accessibility, and user acceptance of the technology [4]. Other research adopted the Bioethical Analysis framework for the ethical consideration and policy recommendations regarding the use of Embodied AI in the "Disability 4.0" era [8]. This method focuses on qualitatively identifying and discussing the social impact, fairness, and regulatory needs of technology application.

3 Results

This study obtained a series of findings regarding the technical foundations, clinical efficacy, and core challenges of Embodied Artificial Intelligence (EmAI) in healthcare, based on a systematic review and synthesis of fifteen publications. These findings, derived from the quantitative and qualitative data extracted on hardware, algorithms, and evaluation strategies, form a comprehensive picture of medical EmAI systems.

The Table 1 below reports the key findings from the systematic review, covering dimensions from technological evolution and clinical performance to ethical challenges.

Table 1. Key findings from the systematic review

Category	Finding Theme	Key Findings/Quantitative Data	Supporting Technology
Clinical Efficacy	AI-Assisted Surgery Performance	Average operative time reduced by 25%; complications decreased by 30%; surgical precision improved by 40%.	MLLMs, 3D Segmentation
User Functionality	Assistive Technology Empowerment	Significant improvements observed in navigation, safety, and social engagement for the visually impaired.	Embodied Perception, VFS Scale
System Architecture	Perception & Decision Mechanism	Lidar/depth cameras enable multimodal fusion; decisions use a self-learning AI cycle; PaLM-E failure detection achieved F1=0.91.	AI Wheelchairs, MLLMs
Technical Challenges	Data Generalization & Robustness	Insufficient 3D spatial consistency and few-shot adaptability; necessitates Federated Learning (FL).	MLLMs/SAM Extensions
Ethical Challenges	Equity & Safety Governance	Focus on the digital divide of "Disability 4.0," responsibility traceability, and collaboration boundaries.	Bioethical Analysis

4 Discussion

This study systematically reviews fifteen publications to analyze the technological foundations, clinical efficacy, and deep-seated challenges of Embodied Intelligence (EmAI) in healthcare. By integrating multimodal large models (MLLMs), World Models, and embodied perception/decision mechanisms, EmAI is facilitating a paradigm shift from traditional "screen intelligence" to "scene intelligence" [1]. This transition extends AI's influence beyond data analysis into real-world physical interaction and clinical operations. This discussion interprets key findings, contextualizes them within broader preceding research, and ultimately deduces the critical pathways for achieving EmAI's generalization and safety in the medical domain.

4.1 Technical Value, Paradigm Shift, and Validation of Clinical Benefits

EmAI's superior performance originates from its capability for physical interaction with environments. Hardware components such as Lidar, depth cameras, and sEMG sensors enable efficient multimodal fusion [5,6], providing the physical basis for constructing accurate "World Models." In advanced surgical robots, MLLMs further this capability by modeling spatio-temporal scene graphs [11,12]. This modeling deepens the system's "understanding of reality" and its ability to predict critical actions, marking a major leap in intelligence from abstract reasoning to embodied manipulation [13].

The study's results provide compelling quantitative support: the review by Wah shows that AI-assisted surgery achieves significant clinical benefits, including a 40% improvement in precision, 30% reduction in complications, and a 25% decrease in operative time [3]. These findings suggest that EmAI algorithms are moving beyond mere auxiliary tools to become intelligent decision-making collaborators that actively enhance clinical outcomes. In assistive technology, Zhang's long-term observational study using the VFS scale validated EmAI’s value, showing significant improvements in navigation, safety, and social engagement for the visually impaired, reinforcing a "human-centered" design philosophy that empowers independence [10].

4.2 Key Technical Challenges for Embodied AI in Medical Settings

The broad application of EmAI fundamentally depends on overcoming two core technical challenges: the generalizability of algorithms and the robustness of decision-making.

4.2.1 Data Generalization, Few-Shot Learning, and Federated Learning

The fragmented, scarce, and restricted nature of medical data severely constrains the data generalization capability of EmAI models. Tian et al. highlight that current 3D segmentation models struggle with 3D spatial consistency and few-shot adaptability [7]. Even MLLM-based models face major technical barriers in domain adaptation from general to specific clinical lesions. To resolve data privacy and fragmentation, Federated Learning (FL) is confirmed as a critical path, aggregating multi-institutional knowledge without sharing raw data, which is essential for enhancing generalization and building robust World Models [2].

4.2.2 Real-Time Decision Robustness and Dynamic World Model Adaptation

In high-dynamic settings, such as surgery and telemedicine, real-time decision robustness is paramount for safety [14]. While autonomous navigation employs a self-learning AI cycle for efficiency [5], existing "World Models" lack sufficient predictive accuracy against real-world unpredictability [15]. MLLMs trained with embodied data, exemplified by PaLM-E’s high-precision failure detection (F1=0.91) [12], show potential for improved robustness. Qiu et al. stress the need for real-time feedback and dynamic adaptability [15]. This resonates with Cañamero’s concept of Embodied Affect [9], requiring the World Model to predict dynamic changes in human behavior and physiological states to ensure the safety and reliability of embodied mechanisms during emergencies.

4.3 Critical Bioethical and Human-Robot Interaction Safety Issues

EmAI's physical presence and autonomous capabilities create more complex ethical and safety challenges than traditional AI.

4.3.1 Ethical and Legal Dilemmas: Responsibility Traceability and Collaboration Boundaries

As system autonomy grows, defining the boundaries of human-machine collaborative decision-making and the clear attribution of responsibility is a core legal and ethical dilemma. Duan proposed an explicit Responsibility Traceability Framework to distinguish the liabilities among the human physician, the AI developer, and the device manufacturer in telemedicine, thereby resolving the responsibility vacuum and fostering clinical trust [14]. Cañamero also advocates for integrating Embodied Affect into design to enhance collaborative safety and comfort, which is crucial for long-term assistive device users [9].

4.3.2 Social Equity and "Disability 4.0" Ethical Considerations

The social ethical core of EmAI centers on achieving equity and accessibility. The "Disability 4.0" bioethical analysis and related research focus on cost, accessibility, and user acceptance [4, 8]. The risk is that if these high-end systems are limited to a privileged few, they will inevitably exacerbate the digital divide in healthcare, particularly for vulnerable populations like the disabled and the elderly. Policy formulation must therefore enforce affordability and accessibility as core design constraints, adhering to the principles of Responsible Research and Innovation (RRI), to ensure that the development of EmAI genuinely promotes social equity and universal benefit.

5 Conclusion

This systematic review successfully mapped the current applications, technological foundations, and future trajectory of Embodied Artificial Intelligence (EmAI) in the healthcare sector. The central conclusion of this work is that EmAI represents an irreversible and necessary leap, fundamentally advancing medical intelligence from mere analytical support to reliable, physically integrated clinical collaboration. By establishing a robust framework that seamlessly links multimodal perception, large-model reasoning, and practical physical action, EmAI lays a critical foundation for the construction of truly autonomous and globally accessible next-generation smart healthcare systems.

References

1. Y. Liu, X. Cao, T. Chen, Y. Jiang, J. You, M. Wu, X. Wang, M. Feng, Y. Jin, J. Chen, From screens to scenes: A survey of embodied AI in healthcare. Inf. Fusion 119, C (2025). https://doi.org/10.1016/j.inffus.2025.103033
2. F. Han, X. Huang, X. Wang, Y. Chen, C. Lu, S. Li, … D. Zhang, Artificial Intelligence in Orthopedic Surgery: Current Applications, Challenges, and Future Directions. MedComm 6(7) (2025). https://doi.org/10.1002/mco2.70260
3. J.N.K. Wah, The rise of robotics and AI-assisted surgery in modern healthcare. J. Robotic Surg. 19, 311 (2025).
4. A.-K. Al-Tamimi, L. Hewitt, D. Cameron, M. Salem, A. Moemeni, Challenges of Integrating Assistive Technologies and Robots with Embodied Intelligence in the Homes of Older People Living with Frailty. Appl. Sci. 15(15), 8415 (2025). https://doi.org/10.3390/app15158415
5. J.F. Leaman, Z. Yang, Y.N. Elglaly, H. La, B. Li, Embodied-AI Wheelchair Framework with Hands-free Interface and Manipulation, in Proc. IEEE Int. Conf. Systems, Man, and Cybernetics (SMC), Prague, Czech Republic (2022), pp. 40–45. doi: 10.1109/SMC53654.2022.9945310
6. X. Wu, J. Liang, Y. Yu, G. Li, G.G. Yen, H. Yu, Embodied Perception Interaction, and Cognition for Wearable Robotics: A Survey. IEEE Trans. Cogn. Dev. Syst. (2024). doi: 10.1109/TCDS.2024.3463194
7. Y. Tian, M. Shi, X. Zhang, et al., Assisting embodied AI: a survey of 3D segmentation models for medical CT images. CCF Trans. Pervasive Comput. Interact. (2025). https://doi.org/10.1007/s42486-025-00203-1
8. F. De Micco, V. Tambone, P. Frati, M. Cingolani, R. Scendoni, Disability 4.0: Bioethical Considerations on the Use of Embodied Artificial Intelligence. Front. Med. 11 (2024). https://doi.org/10.3389/fmed.2024.1437280

9. L. Cañamero, Embodied Affect for Real-World Human-Robot Interaction, in Proc. ACM/IEEE Human-Robot Interaction (2020), pp. 459–460. https://doi.org/10.1145/3319502.3374843
10. Q. Zhang, Y. Zhou, Embodied Intelligence Applications in Assistive Technologies for the Visually Impaired, in Proc. 6th Int. Conf. Robotics, Intelligent Control and Artificial Intelligence (RICAI) (2024), pp. 105–109. https://doi.org/10.1109/RICAI64321.2024.10911190
11. W. Huang, J. Yi, Z. Liu, W. Zhou, MLLMs for Versatile Scene Understanding: Towards Embodied Intelligent Surgical Robots, in Proc. 5th Int. Conf. Artificial Intelligence and Industrial Technology Applications (AIITA), Xi'an, China (2025), pp. 1300–1303. doi: 10.1109/AIITA65135.2025.11047944
12. D. Driess, F. Xia, M.S.M. Sajjadi, et al., Palm-E: An embodied multimodal language model (2023).
13. J. Dai, H. Xu, T. Chen, et al., Artificial intelligence for medicine 2025: Navigating the endless frontier. Innovation Medicine 3, 100120 (2025). https://doi.org/10.59717/j.xinn-med.2025.100120
14. H. Duan, Research on the Ethical Resilience of Embodied Intelligence in Telemedicine: Design of Human-Machine Collaborative Decision-Making Boundary and Responsibility Traceability Framework, Artificial Intelligence and Digital Ethics (ICAIDE) (2025), pp. 12–19. https://doi.org/10.1109/ICAIDE65466.2025.11189534
15. Y. Qiu, X. Chen, X. Wu, Y. Li, P. Xu, K. Jin, X. Shang, P. Chotcomwongse, M. He, D. Shi, Embodied Artificial Intelligence in Ophthalmology. npj Digit. Med. 8(1) (2025). https://doi.org/10.1038/s41746-025-01754-4